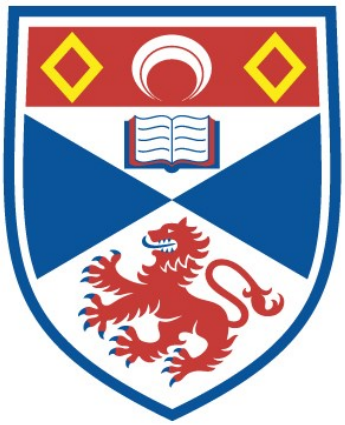
TRANSFORMATIONS IN REGRESSION, ESTIMATION, TESTING AND MODELLING

Imelda Parker

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



1988

TRANSFORMATIONS IN REGRESSION, ESTIMATION, TESTING AND MODELLING

Imelda Parker

A thesis submitted for the Degree of Doctor of Philosophy at the University of St Andrews

Department of Statistics

University of St Andrews

May 1987

DECLARATION

I Imelda Parker hereby certify that this thesis has been composed by myself, that it is a record of my own work, and that it has not been accepted in partial or complete fulfilment of any other degree of professional qualification.

Signed **Date** 13 th May 1987

DECLARATION

I was admitted to the Faculty of Science of the University of St Andrews under Ordinance General No 12 in October 1984 and as a candidate for the degree of Ph.D. in October 1985.

Signed **Date** 13 th May 1987

DECLARATION

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker.

Signed **Date** 13 th May 1987

DECLARATION

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate to the Degree of Ph.D.

Signature of Supervisor **Date** 13th May 1987

T O M Y P A R E N T S

ACKNOWLEDGEMENTS

I would like to thank my Supervisor Mr C D Sinclair for his constant guidance and encouragement. I am indebted to him for many helpful discussions.

Gratitude is due to Professor R M Cormack for the opportunity to pursue this work at the University of St Andrews.

I must also express my appreciation for the excellent typing by Ms Shiela Wilson.

Finally, I am grateful to Philip for his patience and support.

ABSTRACT

Transformation is a powerful tool for model building. In regression the response variable is transformed in order to achieve the usual assumptions of normality, constant variance and additivity of effects. Here the normality assumption is replaced by the Laplace distributional assumption, appropriate when more large errors occur than would be expected if the errors were normally distributed. The parametric model is enlarged to include a transformation parameter and a likelihood procedure is adopted for estimating this parameter simultaneously with other parameters of interest. Diagnostic methods are described for assessing the influence of individual observations on the choice of transformation. Examples are presented.

In distribution methodology the independent responses are transformed in order that a distributional assumption is satisfied for the transformed data. Here the interest is in the family of distributions which are not dependent on an unknown shape parameter. The gamma distribution (known order), with special case the exponential distribution, is a member of this family. An information number approach is proposed for transforming a known distribution to the gamma distribution (known order). The approach provides an insight into the large-sample behaviour of the likelihood procedure considered by Draper and Guttman (1968) for investigating transformations of data which allow the transformed observations to follow a gamma distribution. The

information number approach is illustrated for three examples and the improvement towards the gamma distribution introduced by transformation is measured numerically and graphically.

A graphical procedure is proposed for the general case of investigating transformations of data which allow the transformed observations to follow a distribution dependent on unknown threshold and scale parameters. The procedure is extended to include model testing and estimation for any distribution which with the aid of a power transformation can be put in the simple form of a distribution that is not dependent on an unknown shape parameter. The procedure is based on a ratio, R(y), which is constructed from the power transformation. Also described is a ratio-based technique for estimating the threshold parameter in important parametric models, including the three-parameter Weibull and lognormal distributions. Ratio estimation for the Weibull distribution is assessed and compared with the modified maximum likelihood estimation of Cohen and Whitten (1982) in terms of bias and root mean squared error, by means of a simulation study. The methods are illustrated with several examples and extend naturally to singly Type 1 and Type 2 censored data.

CONTENTS

CHAPTER 1

INTRODUCTION

1.1 TRANSFORMATIONS

Transformation is a powerful tool in regression and distribution theory for converting data into a form that satisfies, at least approximately, the assumptions of a convenient parametric model. An important class of transformations is given by the power transformation family

$$y^{(\lambda)} = \begin{cases} (y^{\lambda} - 1)/\lambda & (\lambda \neq 0) \\ \log y & (\lambda = 0), \end{cases} \qquad (1.1.1)$$

where y represents the data or variable as it arises naturally and is necessarily positive. This continuous family depends on a single parameter λ and all the usual transformations are included: $\lambda = 1$ corresponds to no transformation, $\lambda = -1$ corresponds to the reciprocal transformation and $\lambda = 1/2$ to the square root transformation. When the appropriate λ is unknown the parametric model has one additional unknown parameter λ to be simultaneously estimated with the parameters in the original model.

In regression the usual assumptions behind the linear model are homogeneity of variance, additivity of effects and at least approximate normality of the errors. It may be that a transformation of the response is necessary in order to achieve

these three requirements. Box and Cox (1964) proposed a likelihood approach for selecting transformations of dependent variables in least squares regression. Section 1.2 looks at this likelihood approach and how it could be based on a criterion other than least squares. The implications form the subject of Chapter 2.

In distribution theory it may be necessary to transform data in order to allow the assumption of a distribution to be validly applied to the transformed observations. Section 1.3 presents a preliminary treatment of new statistical methods for checking models and estimating the transformation parameter and other parameters of interest. The methods are developed and discussed in detail in Chapters 3 and 4.

1.2 TRANSFORMATIONS IN REGRESSION

Regression analysis is concerned with the fitting to data of models in which there is a response dependent upon the values of explanatory variables. The models include a statistical error term. It is convenient to assume that the errors have zero mean and constant variance and that the expected value of the response has an additive structure. The relationship between the variables depends upon the values of unknown parameters. The most widely used method of estimation is least squares in which the values of the parameter estimates are chosen to minimise the sum of squared deviations between the observed responses and the predictions from

the model. The additional assumption of normally distributed errors leads naturally to least squares.

If one or more of the assumptions made are in doubt we could contemplate transforming the response. Box and Cox (1964) considered the model

$$y^{(\lambda)} = X\beta + \epsilon,$$

where X is a known $n \times p$ matrix of independent variables with ith row x_i^T, β is a vector of p unknown parameters and ϵ is a vector of normally and independently distributed random variables, each with zero mean and variance σ^2.

The likelihood of the original observations under the proposed model is given by

$$(2\pi\sigma^2)^{-n/2} \exp\left\{-(y^{(\lambda)} - X\beta)^T(y^{(\lambda)} - X\beta)/2\sigma^2\right\}J,$$

where

$$J = \prod_{i=1}^{n} \left|\frac{\partial y_i^{(\lambda)}}{\partial y_i}\right| \qquad (1.2.1)$$

is the Jacobian of the transformation. For fixed λ (1.2.1) is, apart from the Jacobian which is independent of β and σ, the likelihood for a least squares problem with response $y^{(\lambda)}$. The maximum likelihood estimate (MLE) of β, represented by $\hat{\beta}(\lambda)$, is therefore the least squares estimate

$$\hat{\beta}(\lambda) = (X^TX)^{-1}X^Ty^{(\lambda)}. \qquad (1.2.2)$$

The residual sum of squares is

$$RSS(\lambda) = y^{(\lambda)^T}\left\{I - X(X^TX)^{-1}X^T\right\}y^{(\lambda)}, \qquad (1.2.3)$$

where I is the $n \times n$ identity matrix. The MLE of σ^2 is

$$\hat{\sigma}^2(\lambda) = \mathrm{RSS}(\lambda)/n. \qquad (1.2.4)$$

Substitution of the expressions for $\hat{\beta}(\lambda)$ and $\hat{\sigma}^2(\lambda)$ into the logarithm of the likelihood given by (1.2.1) yields, apart from a constant

$$L_{max}(\lambda) = -(n/2)\log\hat{\sigma}^2(\lambda) + (\lambda-1)\sum_{i=1}^{n}\log y_i. \qquad (1.2.5)$$

Plotting $L_{max}(\lambda)$ against λ, the maximising value $\hat{\lambda}$ may be estimated and a $100(1-\alpha)$ per cent confidence interval for λ is found from those values for which

$$2\left[L_{max}(\hat{\lambda}) - L_{max}(\lambda)\right] \leq \chi^2_{1,\alpha}. \qquad (1.2.6)$$

Schlesselman (1971) investigated the scale invariance of this power transformation procedure. He concluded that the procedure is scale invariant if and only if the design matrix X contains the unit vector.

Because the MLE under normality is the least squares estimator, which is known to be nonrobust, the likelihood procedure with the normal distribution assumption is very sensitive to outliers. However, the reason for the presence of too many large residuals may be that the error distribution has longer tails than the normal distribution. For example, suppose the errors ϵ_i, $i = 1,2,\ldots,n$ follow the Laplace distribution

$$f(\epsilon_i) = (2\theta)^{-1}\exp\left\{-|\epsilon_i|/\theta\right\} \qquad -\infty \leq \epsilon_i \leq \infty, \quad \theta > 0. \qquad (1.2.7)$$

The maximum likelihood argument can now be used to obtain estimates based on the criterion of minimising the sum of absolute errors

$$\sum_{i=1}^{n} |\epsilon_i| .$$

Minimum sum of absolute errors (MSAE) regression is a more robust estimation method than least squares, and is less sensitive to changes in the data.

In Chapter 2 these ideas are explored further. A brief introduction to MSAE regression is presented and a test procedure based on standardised residuals is proposed for detecting outliers in simple linear MSAE regression. The power transformation procedure with the Laplace distribution assumption is treated in detail and diagnostic methods for assessing the influence of individual observations on the transformation are described.

1.3 TRANSFORMATIONS IN DISTRIBUTION THEORY

In areas such as reliability and life testing the interest is in analysing random samples of data from some parent distribution. Many of the statistical methods for analysing data are based on the distributional assumption. If the distributional assumption cannot be validly applied to the data, a suitable transformation of the data can sometimes be found that will permit the assumption to be satisfied.

The gamma distribution is an important lifetime model and includes the exponential distribution as a special case. Draper and Guttman (1968) investigated transformations of data which allow the transformed observations to follow a gamma distribution

of known order. They adopted a likelihood procedure for estimating the transformation parameter and the gamma scale parameter.

Chapter 3 introduces an information number approach for transforming a known distribution to the gamma distribution. The approach provides an insight into the large-sample behaviour of the likelihood procedure. The Kullback-Leibler information number is used as a measure of discrepancy between two distributions and facilitates numerical measurement of the improvement to the gamma distribution introduced by transformation. The information number approach also determines the approximate relationships between some important distributions and the gamma distribution.

Statistical procedures in reliability and life testing make use of the relationships that exist between distributions. The procedures are concerned with testing and estimating parametric models for lifetime data. Any results derived in terms of one distribution are easily transferred to the other. Chapter 4 takes this one step further by employing the relationship itself to derive results. A ratio is constructed from the transformation which defines the relationship between a hypothesised distribution and some other distribution. The properties of the ratio for the hypothesised distribution can be used to judge the appropriateness of the model and to supply parameter estimates within the model.

Chapter 5 is devoted to the discussion of the statistical methods developed in earlier chapters with special attention given to areas of further research.

CHAPTER 2

TRANSFORMATIONS, OUTLIERS AND INFLUENTIAL OBSERVATIONS IN MINIMUM SUM OF ABSOLUTE ERRORS REGRESSION

2.1 INTRODUCTION

The most common method of fitting linear models is least squares regression. The usual assumptions behind the linear model include homogeneity of variance, additivity and an error distribution which is symmetric and approximately normal. When more large deviations are observed than would be expected if the errors were normally distributed, a long-tailed error distribution like the Laplace may be more appropriate. However, least squares regression is far from optimal for long-tailed error distributions. Least squares estimates perform poorly and indeed if the variance of the errors is infinite, a situation which has been found to arise in some models, see Huber (1972), the least squares estimates have infinite variance. Furthermore, least squares regression is very sensitive to outliers.

Minimum sum of absolute errors (MSAE) regression gives little weight to outliers and owes its usefulness to its resilience in not being influenced by a few large errors. MSAE regression is a special case of M-estimation and the effect of outliers and leverage points, that is points with leverage values near one, on robust regression estimators is discussed by Krasker

and Welsch (1982) and Hampel et al. (1986). A detailed treatment of the methods and applications of MSAE regression is presented in Bloomfield and Steiger (1983). MSAE regression is optimal when the errors follow the Laplace distribution, that is the MSAE estimates are also the maximum likelihood estimates. Therefore in a situation where outliers are likely to occur, it seems appropriate to consider a linear model with Laplace errors and use an MSAE regression.

It may be necessary to transform the response in order to obtain a model which would satisfy the three requirements of homogeneity of variance, additivity and a Laplace error distribution. Diagnostics such as plots of the residuals versus the fitted values or other variables, can suggest possible transformations. A more objective approach is to use the distributional assumptions concerning the errors and choose the transformation to maximise a criterion function of interest. The chosen criterion is the likelihood of the original observations, used by Box and Cox (1964) in their approach to determining transformations in least squares regression with the assumption of normal errors.

The information for a transformation may depend on one or a few observations. It is therefore important to know whether the evidence for a particular choice of transformation is spread evenly throughout the data or is being unduly influenced by one or more observations. For least squares analyses, Atkinson (1982) and Cook and Wang (1983) have developed diagnostic methods for identifying influential observations. Diagnostics for assessing

the contribution of individual observations to the evidence for a transformation are required for MSAE analyses.

The main stumbling block to applying MSAE methods in the past has been the great computational difficulty involved. However, Charnes et al. (1955) showed that MSAE estimates emerge as solutions to linear programming problems using simplex methods. Barrodale and Roberts (1973) modified the method to save storage and improve efficiency. All MSAE estimates in this article are computed using the NAG subroutine H01ADF which employs the Revised Simplex Method as described by Gavin (1960).

The next section presents some results for MSAE regression. Section 2.3 describes a test for a single outlier in simple MSAE regression. In Section 2.4 the theory is developed for power transformations of the response to find a model which satisfies the requirements of additivity, homogeneity of variance and a Laplace error distribution using MSAE regression. The scale invariance of the power transformation procedure is investigated in Section 2.5. Diagnostic plots for the influence of individual observations on evidence for a transformation are described in Section 2.6 and applied to examples in Section 2.7. The results of least squares and MSAE analyses are compared using probability plots of residuals, specifically the stabilised probability plots of Michael (1983) described in Appendix A.

2.2 MINIMUM SUM OF ABSOLUTE ERRORS REGRESSION

Let the model to be estimated be given by

$$y = X\beta + \epsilon, \qquad (2.2.1)$$

where y is a response variable of n observations, X is a known n×p matrix with ith row X_i^T, β is a vector of p unknown parameters and ϵ is an n×1 vector of randomly distributed errors.

We wish to estimate β by minimising the expression

$$\sum_{i=1}^{n} |y_i - X_i^T\beta|$$

with respect to β. Let $\tilde{\beta}$ represent the MSAE estimator. The MSAE regression can be formulated as a linear programming problem. This is done by writing $y - X\tilde{\beta}$ as the difference between two non-negative variables:

$$y_i - X_i^T\tilde{\beta} = r_i^+ - r_i^-, \qquad r_i^+, r_i^- \geq 0 \qquad (2.2.2)$$

and then minimising

$$\sum_{i=1}^{n} r_i^+ + \sum_{i=1}^{n} r_i^-, \qquad (2.2.3)$$

where the ith components of r^+ and r^- are given by

$$r_i^+ = \begin{cases} y_i - X_i^T\tilde{\beta} & \text{if } y_i - X_i^T\tilde{\beta} \geq 0 \\ 0 & \text{otherwise} \end{cases} \qquad (2.2.4)$$

$$r_i^- = \begin{cases} -(y_i - X_i^T\tilde{\beta}) & \text{if } y_i - X_i^T\tilde{\beta} < 0 \\ 0 & \text{otherwise.} \end{cases} \qquad (2.2.5)$$

Sposito, Smith and McCormick (1978) have shown that where the design matrix X is of full rank the regression hyperplane is determined by a subset of p observations. Consequently the system of equations for the MSAE regression can be written as

$$\begin{bmatrix} y_{(1)} \\ y_{(2)} \end{bmatrix} = \begin{bmatrix} X_{(1)} & 0 \\ X_{(2)} & D \end{bmatrix} \begin{bmatrix} \tilde{\beta} \\ r_{(2)} \end{bmatrix}, \qquad (2.2.6)$$

where, degeneracy aside, subscripts (1) and (2) refer to the observations with zero and non-zero residuals, respectively. This corresponds to reordering the observations so that the first p are those lying on the MSAE regression plane. Accordingly, $X_{(1)}$ is a p×p matrix, $X_{(2)}$ is (n-p) × p in dimension and 0 is a p × (n-p) matrix of zeros. The vector $r_{(2)}$ is made up of the components of the vectors r^+ and r^- and refers to the non-zero residuals. Since these are n-p in number, the dimension of the diagonal matrix D is (n-p) × (n-p) with diagonal entries that are either +1 or -1 depending upon whether $r_i^+ > 0$ or $r_i^- > 0$. It follows from (2.2.6) that

$$y_{(1)} = X_{(1)}\tilde{\beta} \qquad (2.2.7)$$

and

$$\tilde{\beta} = X_{(1)}^{-1} y_{(1)}. \qquad (2.2.8)$$

In Section 2.1 the hope was expressed that the combination of MSAE regression and a Laplace error distribution can sometimes reconcile observations outlying for other analyses, to the MSAE model. Methods for detecting departures from the fitted MSAE model are required. The next section presents a procedure for testing for a single outlier in simple linear MSAE regression.

2.3 TESTING FOR A SINGLE OUTLIER

A procedure of standardising residuals by dividing them by their estimated standard deviations was proposed by Tietjen, Moore and Beckman (1973) for testing for a single outlier in simple linear least squares regression. The same approach is adopted here for detecting a single outlier in simple linear MSAE regression.

Consider the regression model

$$Y = X\beta + \epsilon, \qquad (2.3.1)$$

where ϵ is a vector of independent Laplace random variables with mean zero and constant variance σ^2. The MSAE estimate of β is $X_{(1)}^{-1}y_{(1)}$ where (1) represents the set of defining observations. Expressing $y_{(1)}$ as $I_{(1)}y$, where $I_{(1)}$ is the p×n matrix with rows having a one in the position related to the corresponding defining observation and zeros elsewhere, then the vector of residuals may be expressed as

$$r = (I - XX_{(1)}^{-1}I_{(1)})\epsilon. \qquad (2.3.2)$$

Conditional on the defining set, the residual vector has covariance matrix

$$V(r) = [I + X(X_{(1)}^{T}X_{(1)})^{-1}X^{T})]\sigma^2. \qquad (2.3.3)$$

For simple regression the model is

$$y_i = \beta_1 + \beta_2 x_i + \epsilon_i. \qquad (2.3.4)$$

The estimated standard deviation of r_i is the ith diagonal entry of

$$[I + X(X_{(1)}^{T}X_{(1)})^{-1}X^{T}]\tilde{\sigma}^2,$$

where $\tilde{\sigma} = \sqrt{2}\ \Sigma\ |r_i|/(n-2)$ is also conditional on the defining set.

The ith entry is

$$s_i = \tilde{\sigma}\sqrt{\frac{3}{2} + \frac{(x_i - \bar{x}_{(i)})^2}{\sum_{j\in(i)} (x_j - \bar{x}_{(i)})^2}}, \qquad (2.3.5)$$

where $\bar{x}_{(i)}$ is the arithmetic mean of the defining observations.

It is proposed that

$$R_n = \max |r_i/s_i| \qquad (2.3.6)$$

be used as a test statistic for the rejection of a single outlier in linear MSAE regression, a large value of R_n indicating an outlier. R_n is defined to be zero whenever $r_i = 0$ which includes the case for the defining observations. Assuming that $\tilde{\sigma} > 0$ then $s_i > 0$ for all i and R_n is always defined.

A simulation study was conducted to determine the critical values for R_n. For each sample size n, 20,000 samples were generated. The observations were generated as random variables from the Laplace distribution with zero mean and unit variance. The values of the explanatory variable were chosen to be uniformly spaced on [0,1]. A simple MSAE regression model was fitted to each sample and the values of R_n were calculated. The $100(1-\alpha)$ percentage points of R_n for significance levels $\alpha = 0.1$, 0.05 and 0.01 were calculated and recorded in Table 2.1. A range of sample sizes up to n = 50 was considered.

In order to investigate the "power" of the test statistic R_n, the simulations were repeated for samples of size 10, 25 and 40 and a single outlier from a Laplace distribution with mean μ and unit variance was used to contaminate the sample. The "power" plot of Figure 2.1 summarises the results. It shows the percentage of samples in which the outlier was correctly

Table 2.1 Critical values of R_n for detecting single outlier in simple linear MSAE regression

n	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
8	1.647	1.839	2.193
10	1.826	2.048	2.480
12	1.952	2.196	2.669
14	2.096	2.337	2.870
16	2.217	2.471	3.031
18	2.295	2.566	3.175
20	2.371	2.657	3.264
25	2.546	2.845	3.521
30	2.668	2.985	3.593
35	2.754	3.090	3.834
40	2.861	3.209	3.931
45	2.947	3.282	4.096
50	3.012	3.353	4.119

identified by the test statistic R_n, plotted against μ, the mean of the outlying population, for $\mu = 1(1)\ 10$. The plot shows that

the powers do not always increase with sample size, a feature shared by the "power" plot of Tietjen, Moore and Beckman (1973).

The use of the test statistic R_n is exemplified in the following examples.

Example 2.1 Wormy Fruit Data

Snedecor and Cochran (1967, p.150) give twelve observations on the percentage of wormy fruit for different sizes of apple crop. The data are given in Table 2.2. For simple least squares regression, observation 4 gives rise to a large standardised residual taking the value 2.210. For simple MSAE regression the largest standardised residual corresponds to observation 4 and takes the value 0.722. We have $R_n = 0.722$ and consulting Table 2.1 we see that the 10 percent point of R_n is approximately 1.952. It may be concluded that observation 4 is not outlying for the MSAE regression model.

Example 2.2 Tree Data

The data, taken from Chapman and Demeritt (1936), are given in Table 2.3. There are $n = 27$ observations on diameter at breast height (DBH), measured in inches, of chesnut trees grown on a poor site. The response is DBH and the explanatory variable is age in years. Weisburg (1980) used these data to illustrate the inadequacy of a simple least squares regression model. Observation 24 has a large standardised residual of -2.422 for the

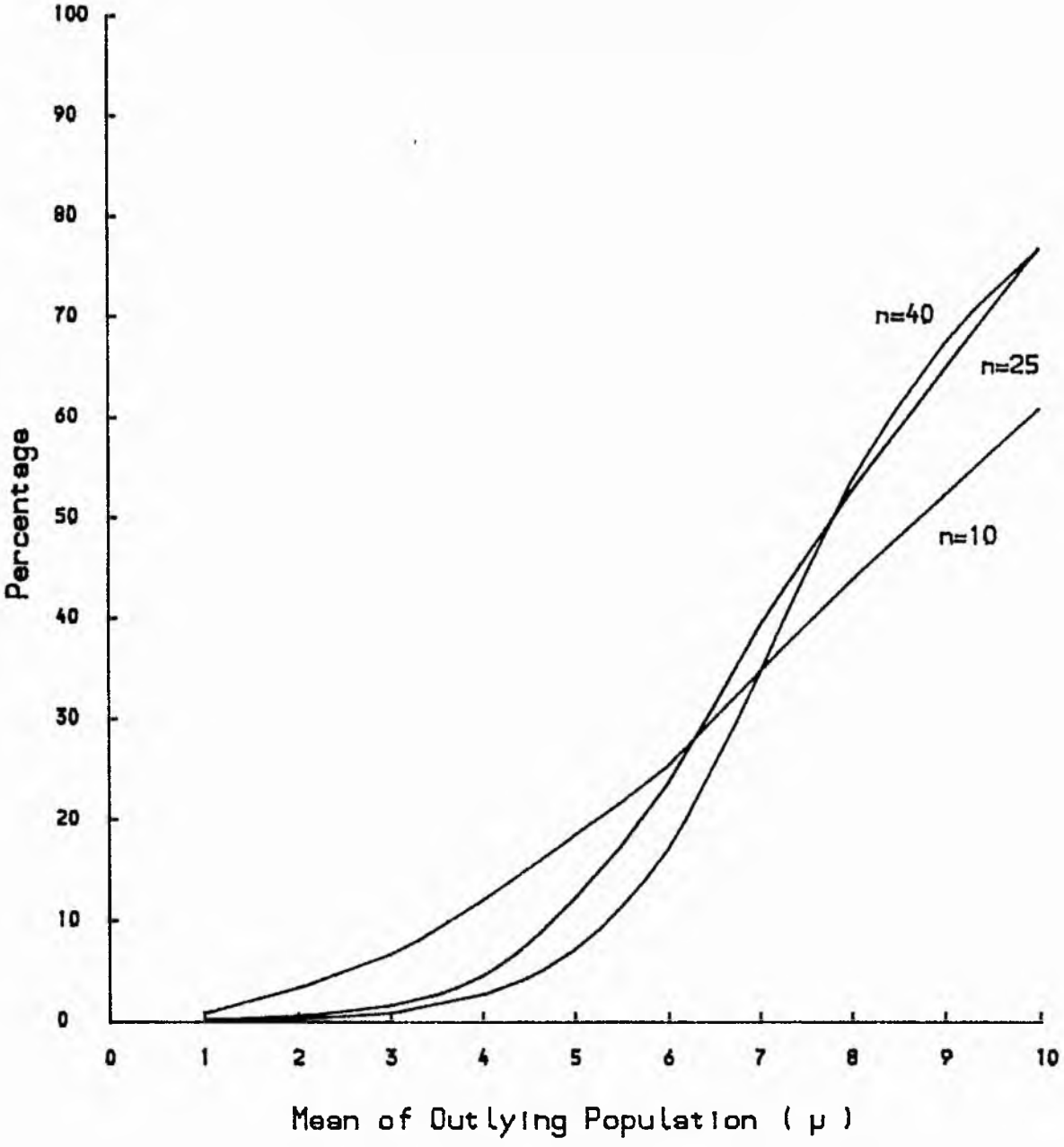


Figure 2.1. Percentage of outliers correctly identified with $\alpha = 0.05$.

Table 2.2 Wormy fruit data

Tree number	Size of crop on tree (hundreds of fruit) X	Percentage of wormy fruit Y
1	8	59
2	6	58
3	11	56
4	22	53
5	14	50
6	17	45
7	18	43
8	24	42
9	19	39
10	23	38
11	26	30
12	40	27

Table 2.3 Tree data

Observation	Age	DBH
1	4	0.8
2	5	0.8
3	8	1.0
4	8	2.0
5	8	3.0
6	10	2.0
7	10	3.5
8	12	4.9
9	13	3.5
10	14	2.5
11	16	4.5
12	18	4.6
13	20	5.5
14	22	5.8
15	23	4.7
16	25	6.5
17	28	6.0
18	29	4.5
19	30	6.0
20	30	7.0
21	33	8.0
22	34	6.5
23	35	7.0
24	38	5.0
25	38	7.0
26	40	7.5
27	42	7.5

least squares model.

For the simple MSAE model, the test statistic $R_n = 1.332$. Entering Table 2.1 with n = 27 we see that the 10 percent point of R_n is approximately 2.6. The conclusion is that the observation is not an outlier for the MSAE model.

These examples clearly show how including a Laplace error distribution and using MSAE regression can accommodate observations outlying for other analyses. The assumptions behind the MSAE linear modelling also included homogeneity of variance and simplicity of structure for the expected value of the response. If these requirements are not satisfied in the original scale of measurement of the response, it may be that there is a transformation of the response which would allow the assumptions to be validly applied. Section 2.4 describes a likelihood procedure for selecting transformations.

2.4 THE BOX-COX TRANSFORMATION

Consider the positive response variable y of n observations and the power transformation family of Box and Cox (1964)

$$y^{(\lambda)} = \begin{cases} (y^{\lambda} - 1)/\lambda & (\lambda \neq 0) \\ \log y & (\lambda = 0) \end{cases} \qquad (2.4.1)$$

It is hypothesised that for some λ

$$y^{(\lambda)} = X\beta + \epsilon, \qquad (2.4.2)$$

where the errors ϵ_i, $i = 1,2,\ldots,n$ follow the Laplace distribution $f(\epsilon_i) = (2\theta)^{-1} \exp\{-|\epsilon_i|/\theta\}$, $\theta > 0$.

The hypothesised model of Box and Cox (1964) included a normal error distribution. They used maximum likelihood estimation to simultaneously estimate λ, β and the constant error variance. However, the maximum likelihood estimator (MLE) is very sensitive to outliers (Carroll 1982). Carroll (1980) proposed a robust method for transforming to achieve approximate normality in a linear model. Carroll and Ruppert (1985) introduced several robust estimation techniques for the Box-Cox transformation model.

The likelihood procedure of Box and Cox (1964) is adapted for the hypothesised model of (2.2). The likelihood of the original observations is

$$(2\theta)^{-n} J \exp\left\{-\sum_{i=1}^{n} |y_i^{(\lambda)} - x_i^T \beta|/\theta\right\}, \qquad (2.4.3)$$

where the Jacobian is

$$J = \prod_{i=1}^{n} \left|\frac{\partial y_i^{(\lambda)}}{\partial y_i}\right|. \qquad (2.4.4)$$

For λ fixed, (2.4.3) is the likelihood corresponding to a standard MSAE problem with response $y^{(\lambda)}$, apart from the constant factor of the Jacobian. The MLE of β for given λ, which we call $\tilde{\beta}(\lambda)$, is therefore the MSAE estimate

$$\tilde{\beta}(\lambda) = X_{(1)}^{-1} y_{(1)}^{(\lambda)}, \qquad (2.4.5)$$

where subscript (1) represents the defining observations. The MLE

of θ is

$$\tilde{\theta}(\lambda) = \frac{1}{n} \sum_{i=1}^{n} |y_i^{(\lambda)} - \underset{\sim}{x}_i^T \tilde{\beta}(\lambda)|. \qquad (2.4.6)$$

Substitution of the MLEs for β and θ in the logarithm of the likelihood yields, apart from a constant,

$$L_{max}(\lambda) = -n \log \left[\sum_{i=1}^{n} |y_i^{(\lambda)} - \underset{\sim}{x}_i^T \tilde{\beta}(\lambda)| \right] + \log J.$$

This partially maximised log-likelihood is a function of λ and depends on the Jacobian J. Working with the normalized transformation

$$z^{(\lambda)} = y^{(\lambda)}/J^{1/n}$$

reduces $L_{max}(\lambda)$ to a simpler, but equivalent, form. The Jacobian of the transformation (2.4.1) is

$$J = \prod_{i=1}^{n} y_i^{\lambda-1}.$$

The normalized power transformation may be expressed as

$$z^{(\lambda)} = \begin{cases} (y^{\lambda} - 1)/\lambda \dot{y}^{\lambda-1} & (\lambda \neq 0) \\ \dot{y} \log y & (\lambda = 0), \end{cases} \qquad (2.4.7)$$

where $\dot{y}$ is the geometric mean of the observations. Except for a constant, the partially maximised log-likelihood can now be written as

$$L_{max}(\lambda) = -n \log \left[\sum_{i=1}^{n} |z_i^{(\lambda)} - \underset{\sim}{x}_i^T \underset{\sim}{x}_{(i)}^{-1} \underset{\sim}{z}_{(i)}^{(\lambda)}| \right] \qquad (2.4.8)$$

Plotting $L_{max}(\lambda)$ against λ, the maximising value $\tilde{\lambda}$ may be estimated. An approximate $100(1-\alpha)$ per cent confidence interval for λ is found from those values for which

$$2[L_{max}(\tilde{\lambda}) - L_{max}(\lambda)] \leq \chi^2_{1,\alpha}. \qquad (2.4.9)$$

Note that (2.4.9) is conditional on θ and β taking the values of their MLEs. The scale invariance of the selection of the transformation parameter is investigated in Section 2.5.

The estimators introduced above are not robust in the sense of bounded influence, but estimators that bound the influence function may be constructed using the method described by Ruppert and Carroll (1985).

An example concludes this section.

Example 2.3 Stack-Loss Data

In this often-analysed data set there are 21 observations on losses of ammonia from an oxidation plant and three explanatory variables. The data, taken from Brownlee (1965, p.454), are given in Table 2.4. Some analyses lead to the dropping of the third explanatory variable, acid concentration. Robust analyses, such as that of Andrews (1974), have identified observations 1, 3, 4, 21 and possibly 2 as outliers.

One reason for the presence of so many outliers may be that the errors are not normally distributed. A long-tailed error distribution like the Laplace combined with MSAE regression may yield a better fit of the data to the model. An exponential probability plot of the absolute residuals (Figure 2.2a) shows no major departures from a straight line. The defining observations

Table 2.4 Stack-loss data

Observation	Air flow	Cooling water inlet temperature	Acid concentration	Stack loss
1	80	27	89	42
2	80	27	88	37
3	75	25	90	37
4	62	24	87	28
5	62	22	87	18
6	62	23	87	18
7	62	24	93	19
8	62	24	93	20
9	58	23	87	15
10	58	18	80	14
11	58	18	89	14
12	58	17	88	13
13	58	18	82	11
14	58	19	93	12
15	50	18	89	8
16	50	18	86	7
17	50	19	72	8
18	50	19	79	8
19	50	20	80	9
20	56	20	82	15
21	70	20	91	15

have zero residuals and are treated as a single observation when plotting Figure 2.2a. The plot demonstrates that this approach goes a long way towards accommodating observations 1, 3, 4 and 21 (with observation 2 among the defining observations). The residuals might still be considered large and a transformation of the data could prove worthwhile.

Atkinson (1982) transformed the response for the least squares analysis. He calculated that the MLE of λ is $\hat{\lambda} = 0.30$. The robust estimators of Carroll and Ruppert (1985) agree that $\lambda = 1/2$ is reasonable.

The plot of the log-likelihood $L_{max}(\lambda)$ against λ for the MSAE analysis is shown in Figure 2.2b. The log-likelihood is a maximum at $\tilde{\lambda} = 0.42$. The approximate 95 per cent confidence interval for λ defined by (2.4.9) covers 0.05 to 0.69. The square root transformation again appears appropriate. MSAE regression in the square root scale leads to dropping the explanatory variable, acid concentration.

The normal probability plot of the residuals for the transformed model and the least squares analysis is shown in Figure 2.3a. It exhibits systematic departure from a straight line with the residuals for observations 4 and 21 appreciably distanced, illustrating that the least squares analysis results in a poor fit of the data to the model. Figure 2.3b shows the exponential plot of the absolute residuals for the transformed model and the MSAE analysis. It is a decided improvement on Figure 2.3a. Observations 4 and 21 have the largest residuals but all observations appear to be adequately modelled.

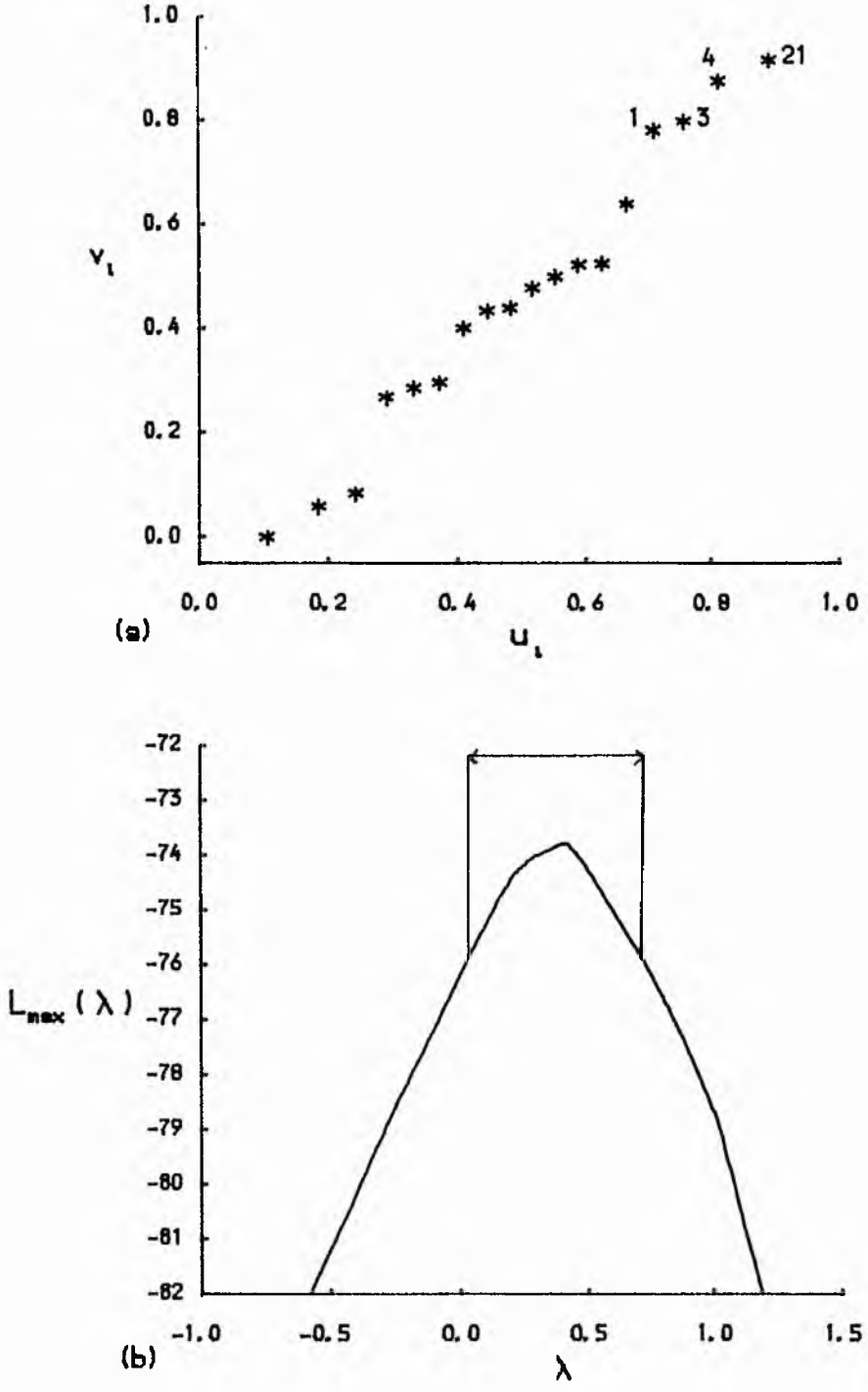


Figure 2.2. Stack-Loss Data. MSAE Analysis:
(a) Exponential plot of absolute residuals
(b) Plot of log-likelihood with 95% confidence interval for λ indicated.

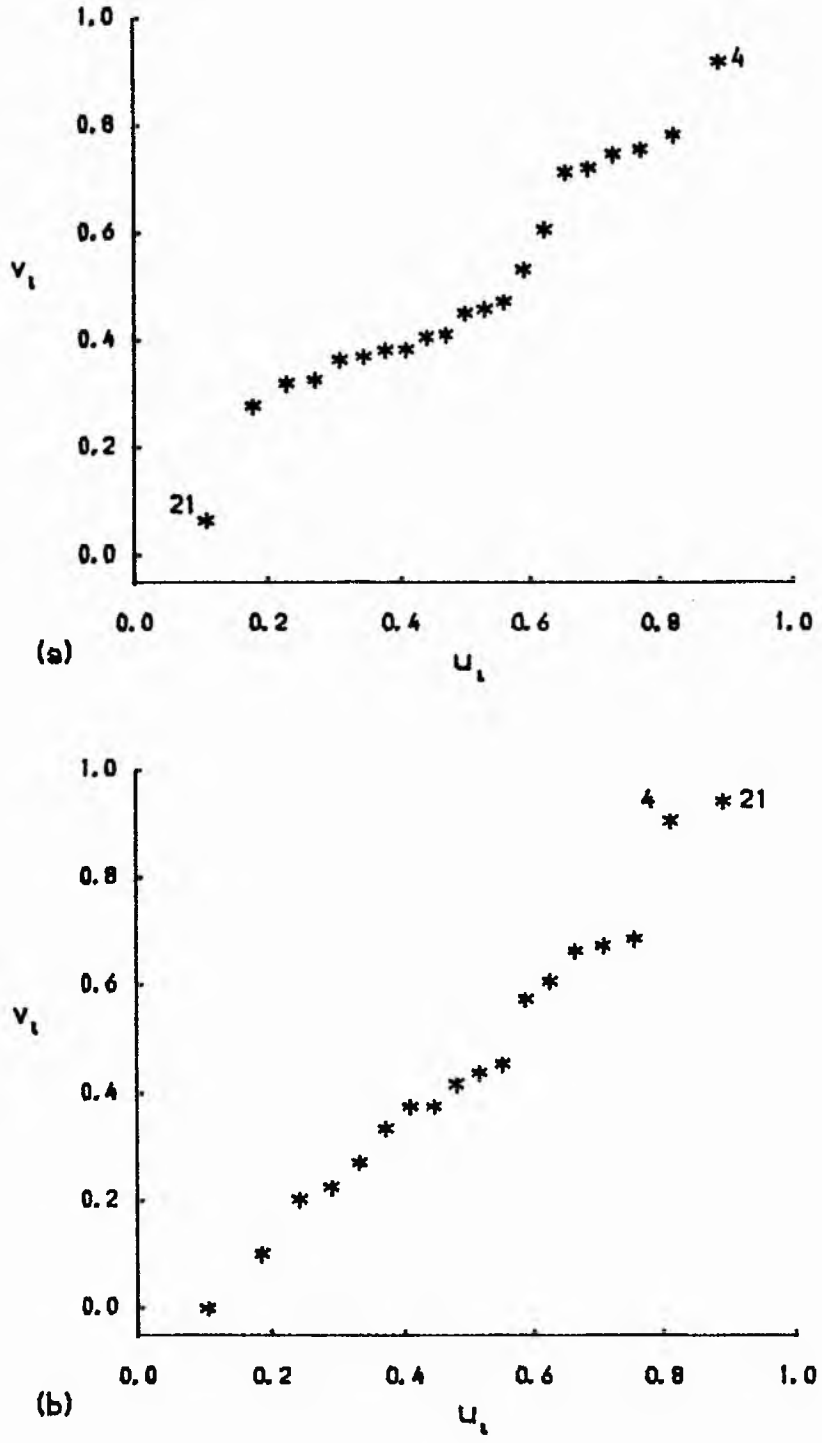


Figure 2.3. Stack-Loss Data. (a) Least Squares Analysis: Normal plot of residuals for transformed model ($\lambda = 1/2$) (b) MSAE Analysis: Exponential plot of absolute residuals for transformed model ($\lambda = 1/2$).

2.5 SCALE INVARIANCE

This section investigates the scale invariance of the likelihood procedure for selecting the transformation parameter when the errors are Laplace. The log-likelihood as a function of λ reads

$$L_{max}(\lambda) = -\, n \log \left[\sum_{i=1}^{n} \left| z_i^{(\lambda)} - X_i^T X_{(i)}^{-1} z_{(i)}^{(\lambda)} \right| \right]. \qquad (2.5.1)$$

Under the rescaling $y \xrightarrow{\omega} \omega y$

$$y^{(\lambda)} \xrightarrow{\omega} \omega^{\lambda} y^{(\lambda)} + (\omega^{\lambda} - 1)/\lambda$$

and

$$J^{1/n} \xrightarrow{\omega} \omega^{\lambda} J^{1/n}.$$

Consequently

$$z^{(\lambda)} \xrightarrow{\omega} z^{(\lambda)} + s \qquad (2.5.2)$$

where $s = \lambda^{-1} (1 - \omega^{-\lambda}) J^{1/n}$.

Suppose the design matrix X allows for removal of an additive constant, that is a column (the first column say) of X consists entirely of ones. Then the first column of X^{-1} sums to one with the remaining columns summing to zero, that is $X^{-1} 1_n = e$ where 1_n is a $n \times 1$ vector of ones and e is the $n \times 1$ vector with first entry one and zeros elsewhere. This result also applies to $X_{(i)}$ as a submatrix of X. It follows that $X_{(i)}^{-1} 1_n = e$ and $X_i^T X_{(i)}^{-1} 1_n = 1$ for $i = 1, \ldots, n$.

But under the rescaling

$$z_i^{(\lambda)} - X_i^T X_{(i)}^{-1} z_{(i)}^{(\lambda)} \xrightarrow{\omega} z_i^{(\lambda)} - X_i^T X_{(i)}^{-1} z_{(i)}^{(\lambda)} + s(1 - X_i^T X_{(i)}^{-1} 1_n) \qquad (2.5.3)$$

for $i = 1,\ldots,n$.

The term in parenthesis on the right-hand side vanishes if and only if X satisfies the above condition and contains the unit vector.

Thus

$$L_{max}(\lambda) \xrightarrow{\omega} L_{max}(\lambda)$$

and the selection of the transformation parameter λ is scale invariant only if X allows for the removal of an additive constant. This agrees with Schlesselman's (1971) conclusion regarding the scale invariance of the likelihood procedure under normality.

2.6 INFLUENCE DIAGNOSTICS

Because the MSAE regression is completely determined by a subset of the observations, influence diagnostics similar to those employed in least squares regression are not readily available. These defining observations have zero residuals and are influential, but not in the usual context. The concept of a subset of defining observations is unique to MSAE regression and no comparison can be drawn with least squares regression. The defining observations are set aside with their own special brand of influence and the influence of the remaining observations on choice of transformation is considered.

Degeneracy aside, only the defining observations have zero residuals and do not contribute to the log-likelihood which may be reexpressed as

$$- n \log \left[\sum_{i \in (2)} \left| z_i^{(\lambda)} - x_i^T x_{(i)}^{-1} z_{(i)}^{(\lambda)} \right| \right], \qquad (2.6.1)$$

where set (2) is the set of observations with non-zero residuals. The slope of the log-likelihood is

$$S(\lambda) = \frac{- n \sum_{i \in (2)} \operatorname{sgn} r_i(\lambda)(w_i(\lambda) - x_i^T x_{(i)}^{-1} w_{(i)}(\lambda))}{\sum_{i \in (2)} |r_i(\lambda)|}, \qquad (2.6.2)$$

where $r(\lambda)$ is the residual vector for the transformed model and $w(\lambda) = \partial z^{(\lambda)}/\partial\lambda$. For the optimal value of the transformation parameter, $\tilde{\lambda}$,

$$S(\tilde{\lambda}) = 0 \qquad (2.6.3)$$

For the hypothesised transformation parameter value λ°, the statistic $S(\lambda^\circ)$ is called the efficient score (Cox and Hinkley, 1978). The efficient score $S(1)$ is a measure of the need for a transformation.

The quantity $\operatorname{sgn} r_i(\lambda^\circ)(w_i(\lambda^\circ) - x_i^T x_{(i)}^{-1} w_{(i)}(\lambda^\circ))$ is a measure of the contribution of the i^{th} observation to the slope of the log-likelihood at λ°. This interpretation provides an analysis of the influence of individual observations on the transformation. The hypothesis of no transformation, that is $\lambda^\circ = 1$, is of particular interest. Sometimes choosing between an untransformed and logged response is a factor when exploring models. An index plot of the influence measure

$$T_i = \operatorname{sgn} r_i(1)(w_i(1) - x_i^T x_{(i)}^{-1} w_{(i)}(1)) \qquad (2.6.4)$$

provides an assessment of the influence of individual observations on the need for a transformation. Large positive (negative) values for T_i are associated with observations which are

influential in determining (denying) the need for a transformation.

An index plot of the MLE of λ after deleting the i^{th} observation, $\tilde{\lambda}^{(i)}$ say, can also provide a useful diagnostic guide to the effect of individual observations on the transformation. Atkinson's (1982) diagnostic plots fail to identify observations influential for a transformation if the observation is also a leverage point. When such a situation arises, Atkinson (1985) uses index plots of deletion estimates of λ to assess the effect of observations on the transformation.

The diagnostic methods described above are exemplified in the examples of the next section.

2.7 EXAMPLES

Example 2.3 (continued) Stack-Loss Data

For his least squares analysis, Atkinson (1982) calculated the MLE of λ to be 0.30. He investigated the contribution of individual observations to the evidence of the need for a transformation. His diagnostic plots identify observation 21 as the most influential. If observation 21 is excluded, then the MLE is $\hat{\lambda} = 0.48$. The robust estimators and diagnostics of Ruppert and Carroll (1985) identify observations 2, 3, 4 and 21 as potential troublemakers.

For the MSAE analysis a large non-zero value of -17.91 for the efficient score S(1) indicates that a transformation of the response is required. The MLE is $\tilde{\lambda} = 0.42$. The index plot of the influence measure T_i (Figure 2.4a) shows that observations 11 and 12 are most influential in denying the need for a transformation. That the evidence for a transformation does not depend crucially on one or more observations is shown by the index plot of $\tilde{\lambda}^{(i)}$, Figure 2.4b. The values are joined by a continuous line and fluctuate around $\tilde{\lambda} = 0.42$ which is represented by a broken line. In particular, deleting observation 21 leaves the MLE unchanged at the value 0.42. Observation 21 has a cancelling effect on observations 11 and 12; deletion of the combinations (11,21), (12,21) and (11,12,21) cause no change to the MLE of λ.

In this example the conclusion obtained using the MSAE analysis is very different from those obtained using earlier analyses. The conclusion is that all 21 observations can be adequately modelled by a first-order model in the first two explanatory variables and the square root scale, using MSAE regression and a Laplace error distribution. In addition, diagnostic plots show no over-riding influence of one or more observations on the choice of transformation.

Example 2.4 Salinity Data

The data with 28 observations on the salinity of water with three regressor variables, salinity lagged two weeks, trend and water flow is from Ruppert and Carroll (1980). The data are given in

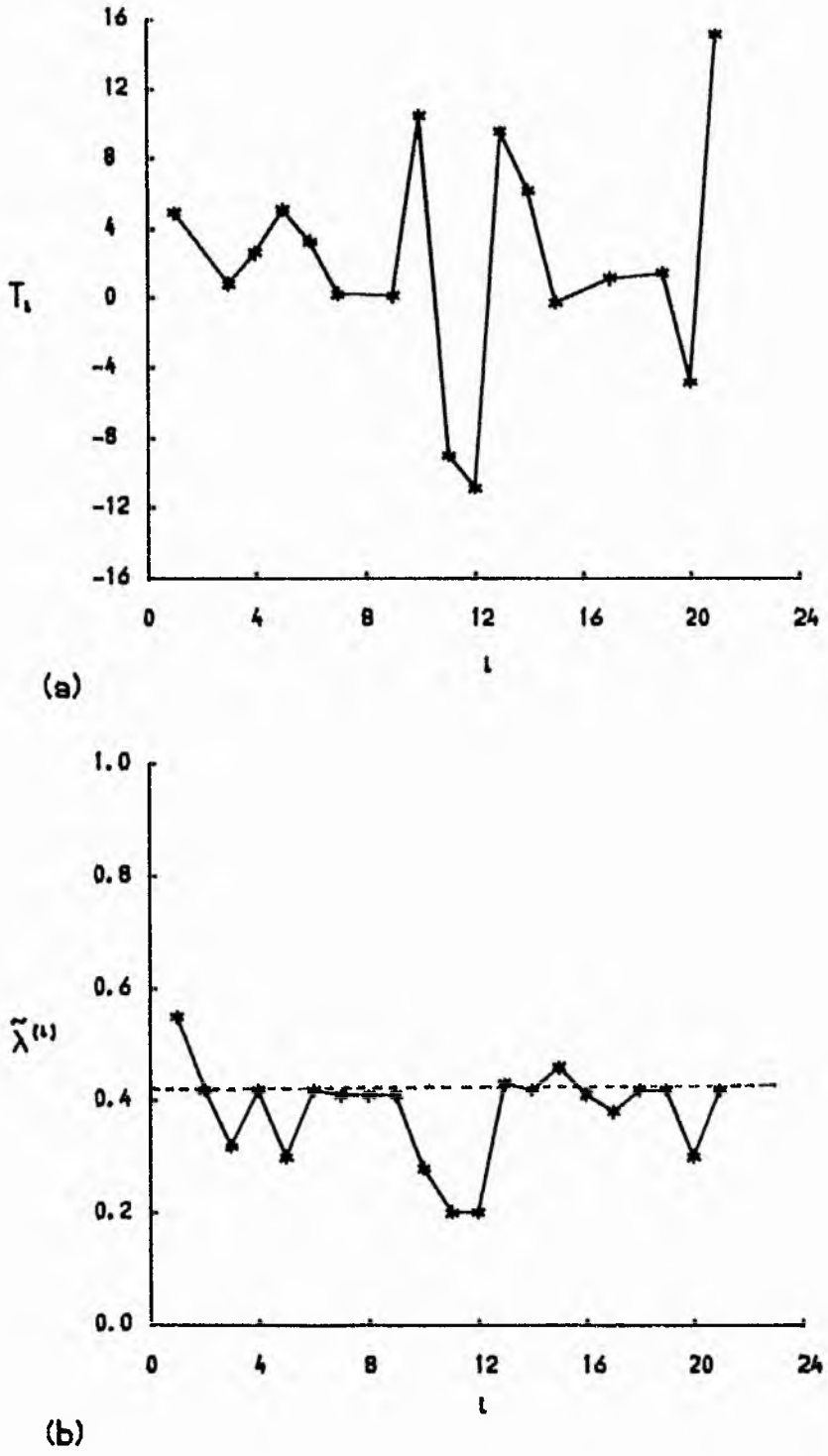


Figure 2.4. Stack-Loss Data. (a) Index plot of influence measure T_i (b) Index plot of deletion MLEs.

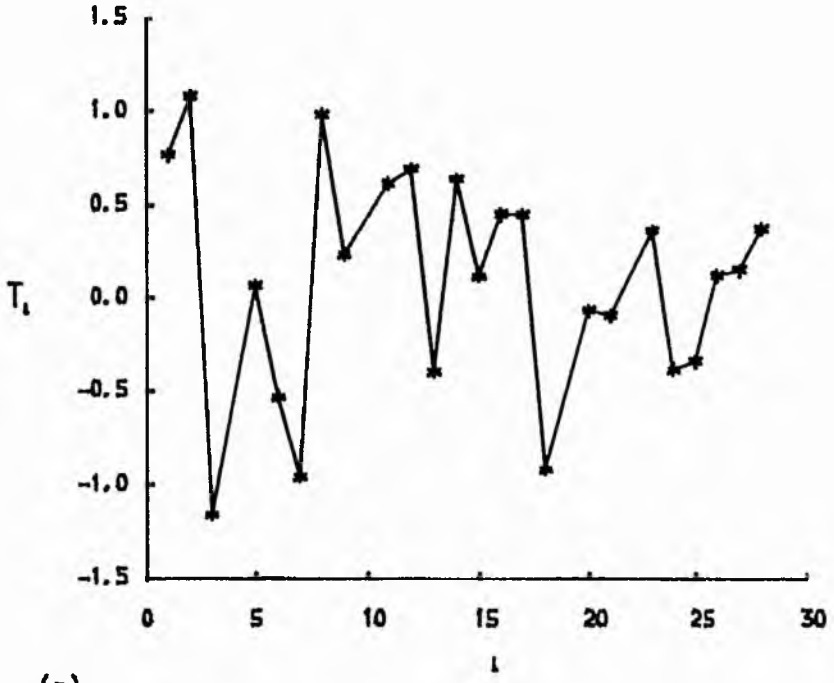
Table 2.5. Atkinson (1983) considered that there was something suspect about the value of the explanatory variable, water flow, corresponding to observation 16 and he 'corrected' it accordingly. In his least squares analysis he calculated the MLE of λ to be $\hat{\lambda} = 0.46$. His diagnostic plots reveal that observation 3 is very influential in denying the need for a transformation. If observation 3 is deleted the MLE is -0.15 and Atkinson's diagnostic plots now show that the evidence for the log transformation is not being particularly influenced by any one observation.

Carroll and Ruppert (1985) considered the salinity data to illustrate their robust estimation for the transformation parameter. They set water flow equal to 26.0 whenever it exceeded 26.0. Their diagnostics indicate the possibility that the effect of observation 5 on estimating λ is masked by observations 3 and 16. This is confirmed by Table 2.6 where, for the least squares analysis, the MLEs of λ with case 'correction' and deletion are arrayed.

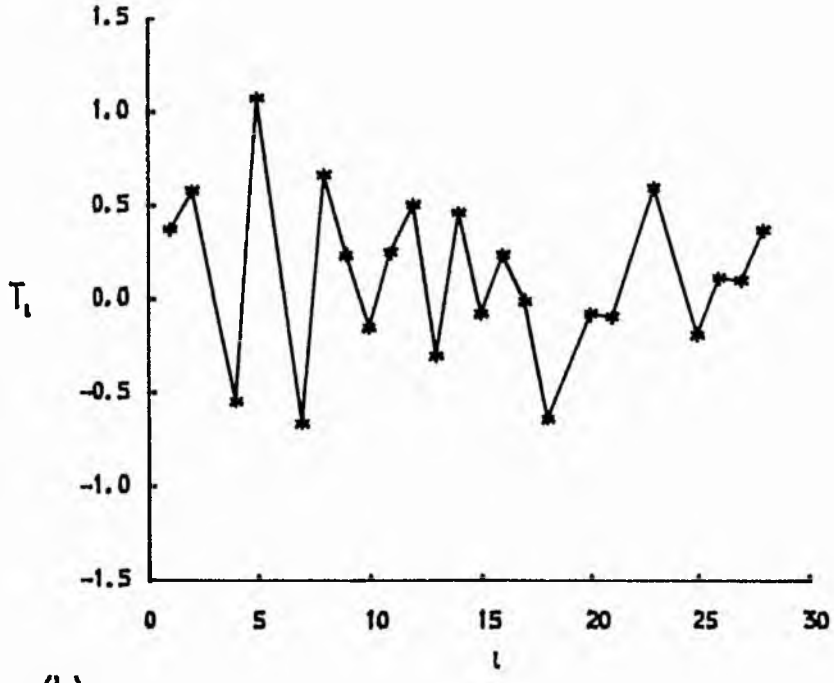
For the MSAE analysis, the value of water flow corresponding to observation 16 is modified following the lead of Atkinson. The MLE of λ is $\tilde{\lambda} = 0.34$. The index plot of influence measure T_i in Figure 2.5a shows that observation 3 is most influential in denying the need for a transformation. If observation 3 is deleted the evidence for a transformation as measured by the efficient score S(1) increases from -3.07 to -4.08 and the MLE is $\tilde{\lambda} = 0.03$. The index plot of influence measure T_i is shown in Figure 2.5b. It reveals that deletion of observation 3 causes

Table 2.5 Salinity data

Observation	Salinity	Lagged Salinity	Trend	Water flow	Year
1	7.6	8.2	4	23.005	72
2	7.7	7.6	5	23.873	
3	4.3	4.6	0	26.417	73
4	5.9	4.3	1	24.868	
5	5.0	5.9	2	29.895	
6	6.5	5.0	3	24.200	
7	8.3	6.5	4	23.215	
8	8.2	8.3	5	21.862	
9	13.2	10.1	0	22.274	74
10	12.6	13.2	1	23.830	
11	10.4	12.6	2	25.144	
12	10.8	10.4	3	22.430	
13	13.1	10.8	4	21.785	
14	12.3	13.1	5	22.380	
15	10.4	13.3	0	23.927	75
16	10.5	10.4	1	33.443	
17	7.7	10.5	2	24.859	
18	9.5	7.7	3	22.686	
19	12.0	10.0	0	21.789	76
20	12.6	12.0	1	22.041	
21	13.6	12.1	4	21.033	
22	14.1	13.6	5	21.005	
23	13.5	15.0	0	25.865	77
24	11.5	13.5	1	26.290	
25	12.0	11.5	2	22.932	
26	13.0	12.0	3	21.313	
27	14.1	13.0	4	20.769	
28	15.1	14.1	5	21.393	



(a)



(b)

Figure 2.5. Salinity Data corrected. (a) Index plot of influence measure T_t (b) Observation 3 deleted: Index plot of influence measure T_t .

observation 5 to become most influential in determining the need for a transformation. After deleting observation 5 the evidence for a transformation as measured by the efficient score S(1) is reduced to −3.37 and the MLE becomes 0.33. Figure 2.5 provides powerful diagnostic plots for the masking of the effect of observations 5 by observations 3 and 16.

Table 2.6 Salinity Data: Maximum likelihood estimates of λ for least squares and MSAE analyses

Data	Least squares $\hat{\lambda}$	MSAE $\tilde{\lambda}$
All observations	0.97	0.44
Observation 16 'corrected'	0.46	0.34
Observation 16 'corrected' Observation 3 'deleted'	−0.15	0.03
Observation 16 'corrected' Observations 3,5 'deleted'	0.37	0.33

This example illustrates that an MSAE analysis can provide useful information to the data analyst. The results for the MSAE analysis are summarised in Table 2.6 where the MLEs of λ with case 'correction' and deletion are recorded. The table shows that for the MSAE analysis the MLE of λ is more robust than the MLE under normality for least squares regression.

2.8 SUMMARY

In many applications of least squares regression there may be reason to believe that the errors are drawn from a distribution that is symmetric, but has longer tails than the normal distribution. It may be that the errors follow the Laplace distribution. Then application of the maximum likelihood principle for estimating parameters would involve minimising the sum of absolute errors, that is MSAE regression.

The purpose of this chapter is to investigate this combination of Laplace errors and MSAE regression. A test procedure is described for detecting a single outlier in simple MSAE regression. The test statistic is the maximum standardised residual and the critical values are determined using a simulation study. A likelihood procedure is proposed for transforming response variables in MSAE regression in order to achieve a model with Laplace errors as well as the usual desiderata of constancy of variance and additivity of effects. Also described are diagnostic methods for identifying cases that influence the choice of a transformation for the response variable.

Application to examples illustrates that this approach can sometimes produce an improvement on the model fitted to the data by previous methods as regards outliers and conformity of residuals to hypothesised distribution. The approach can also contribute useful information on the effects of individual cases.

CHAPTER 3

TRANSFORMING TO THE GAMMA DISTRIBUTION

3.1 INTRODUCTION

The gamma distribution with probability density function (pdf)

$$g_{m,\theta}(t) = \frac{t^{m-1}e^{-t/\theta}}{\Gamma(m)\,\theta^{m}} \qquad 0 \leq x \leq \infty, \quad \theta > 0, \quad m > 0 \qquad (3.1.1)$$

has found application in meteorology, inventory theory, insurance risk theory, economics and queuing theory. In life-testing the gamma distribution is a useful lifetime model for a system if it initially experiences wear-out, reaching a stable state of repair as time passes so that a constant hazard function would then apply. The gamma distribution includes the exponential distribution, itself a widely used model in life-testing, as a special case (m=1).

Suppose we have a sample of data which we would like to treat as following a gamma distribution of known order m. If the assumption is not valid in the original scale of measurement of the observations, it may be that there is a transformation which when applied to the data would allow the transformed data to follow the gamma distribution.

Hernandez and Johnson (1980) proposed a method based on the Kullback-Leibler information number (see Kullback 1968) for transforming a random variable with known distribution to near normality. A similar information number approach is adopted for transforming a known distribution to an approximate gamma distribution of known order. The information number is employed as a measure of discrepancy between two probability distributions.

DEFINITION: For two absolutely continuous pdf's f_1 and f_2 the Kullback-Leibler information number is defined as

$$I[f_1;f_2] = \int f_1(t) \log \left[\frac{f_1(t)}{f_2(t)}\right] dt. \qquad (3.1.2)$$

A parametric family of power transformations is considered. The transformation parameter and the gamma scale parameter are selected so that the information number between the true density of the transformed variable and the gamma distribution of known order is a minimum.

An alternative measure of discrepancy is provided by the L^1 norm of $f_1 - f_2$, that is $\int |f_1(t) - f_2(t)|dt$, but the modulus of a function is analytically awkward. In addition, the L^1 norm gives equal weight to equal differences between f_1 and f_2 and does not take the magnitude of the smaller density into account. The L^2 norm of $\sqrt{f_1} - \sqrt{f_2}$, that is $\int (\sqrt{f_1(t)} - \sqrt{f_2(t)})^2 dt$, and the information number are superior on both counts. However, using

the information number as a measure of discrepancy has a major advantage when transforming data from a known distribution to follow the gamma distribution of known order. The parameter values that minimise the information number are the limiting values of the estimates obtained using a likelihood procedure (Draper and Guttman, 1968).

The information number approach for transforming known distributions to the gamma distribution of known order is presented in Section 3.2 and applied to families of pdf's in Section 3.3. In Section 3.4 the likelihood procedure for transforming data to follow a gamma distribution of known order is described and its large-sample behaviour investigated.

3.2 THE INFORMATION NUMBER APPROACH

Consider a positive random variable Y with pdf f(.) and the simple power transformation family

$$z = \begin{cases} y^{\lambda}, & \lambda \neq 0 \\ \log y, & \lambda = 0. \end{cases} \qquad (3.2.1)$$

Let $f_{\lambda}(.)$ be the pdf of the transformed variable z. The objective is to minimise $I[f_{\lambda}; g_{m,\theta}]$ for suitable choices of θ and λ, given m. For λ fixed, the minimising value of θ is

$$\theta^*(\lambda) = \begin{cases} \frac{1}{m} E_f(y^{\lambda}), & \lambda \neq 0 \\ \frac{1}{m} E_f(\log y), & \lambda = 0. \end{cases} \qquad (3.2.2)$$

For this choice of θ the information number as a function of λ reads

$$K(\lambda) = \begin{cases} m(1-\log m) + \log[\Gamma(m)] + (1-m\lambda)E_f(\log y) \\ \quad + m\log[E_f(y^\lambda)] + E_f\{\log[f(y)]\} - \log\lambda, & \lambda \neq 0 \\ m(1-\log m) + \log[\Gamma(m)] + E_f(\log y) + m\log[E_f(\log y)] \\ \quad + E_f(\log[f(y)]) - (m-1)E_f[\log(\log y)], & \lambda = 0. \end{cases} \qquad (3.2.3)$$

The expressions $\theta^*(\lambda)$ and $K(\lambda)$ are derived in Appendix B. The function $K(.)$ is plotted against λ and the minimising value, λ^*, estimated from the plot.

Improvement towards the gamma distribution can be assessed numerically and graphically on comparison of

(i) the information numbers

$$I[f_{\lambda^*}; g_{m,\theta^*}] \quad \text{and} \quad I[f; g_{m,\theta_Y}],$$

where $\theta^* = \theta^*(\lambda^*)$, $\theta_Y = E_f(Y)/m$ and $I[f; g_{m,\theta_Y}]$ is a measure of how far the original pdf $f(.)$ is from a gamma pdf of order m,

(ii) plots of $f_{\lambda^*}(.)$ and $g_{m,\theta^*}(.)$.

3.3 EXAMPLES

The lognormal, Weibull and Pareto families of pdf's lend themselves readily as examples to the information number approach

for transforming from a known distribution to an approximate gamma distribution of known order.

Example 3.1 Lognormal family

The lognormal distribution has pdf

$$f(y) = \frac{1}{\sqrt{2\pi}\,\sigma y} \exp\left\{-\frac{1}{2\sigma^2}(\log y - \mu)^2\right\} \qquad y > 0, \quad |\mu| < \infty, \ \sigma > 0. \tag{3.3.1}$$

The Kullback-Leibler information number as a function of λ reads

$$K(\lambda) = m(1 - \log m) + \log[\Gamma(m)] - \tfrac{1}{2}[1 + \log(2\pi)] + \tfrac{1}{2}m\lambda^2\sigma^2 - \log(\sigma\lambda). \tag{3.3.2}$$

The minimising values of the transformation parameter and the gamma scale parameter are

$$\lambda^* = \frac{1}{\sqrt{m}\,\sigma} \tag{3.3.3}$$

and

$$\theta^* = \frac{1}{m}\exp\left[\frac{\mu}{\sqrt{m}\,\sigma} + \frac{1}{2m}\right] \tag{3.3.4}$$

respectively. Note that $K(\lambda)$, λ^* and θ^* are independent of the scale parameter e^{μ} and for fixed m, λ^* decreases with increasing σ.

The minimised information number

$$K(\lambda^*) = m(1 - \log m) + \log[\Gamma(m)] + \tfrac{1}{2}\log(m/2\pi) \tag{3.3.5}$$

depends only on m and decreases with increasing m.

The power transformation of a lognormal variable is also lognormal and $\lambda^* = 1$ if and only if the lognormal shape parameter σ takes the value $1/\sqrt{m}$. Therefore, among all possible lognormal distributions, with the Kullback-Leibler information number used

as a measure of discrepancy, the lognormal distribution with $\sigma = 1/\sqrt{m}$ is 'closest' to the gamma distribution of order m. The gamma scale parameter is $\theta^* = 1/m \exp(\mu + 1/2m)$.

For the case of transforming to the gamma distribution of order 4, Figure 3.1 shows that the plots of f_{λ^*} and g_{4,θ^*} for $\sigma = 0.5$ and $\mu = 0$, 1 and 2 are quite close. The plots of f_{λ^*} and g_{9,θ^*} for transforming to the gamma distribution of order 9 and for $\sigma = 1/3$ and $\mu = 0$, 1 and 2, are closer still. The plots are displayed in Figure 3.2.

Example 3.2 Weibull family

The Weibull distribution has pdf of the form

$$f(y) = \frac{\beta}{\alpha}\left[\frac{y}{\alpha}\right]^{\beta-1} \exp\left\{-\left[\frac{y}{\alpha}\right]^{\beta}\right\} \qquad y > 0, \quad \alpha > 0, \quad \beta > 0. \tag{3.3.6}$$

It is readily shown that

$$K(\lambda) = m(1 - \log m) + \log[\Gamma(m)] + \log\left[\frac{\beta}{\lambda}\right] + \left[1 - m\,\frac{\lambda}{\beta}\right]\gamma + m \log\left[\Gamma\left[1 + \frac{\lambda}{\beta}\right]\right] - 1, \tag{3.3.7}$$

where $\gamma \simeq -0.57722$ is Euler's constant and the minimising gamma scale parameter as a function of λ reads

$$\beta^*(\lambda) = \frac{\alpha^{\lambda}}{m}\,\Gamma\left[1 + \frac{\lambda}{\beta}\right]. \tag{3.3.8}$$

Expressions (3.3.7) and (3.3.8) are both independent of the scale parameter α. The approximation

$$\Gamma(1+k) \approx 1 - 0.5748646k + 0.9512363k^2 - 0.6998588k^3 + 0.4245549k^4 - 0.1010678k^5$$

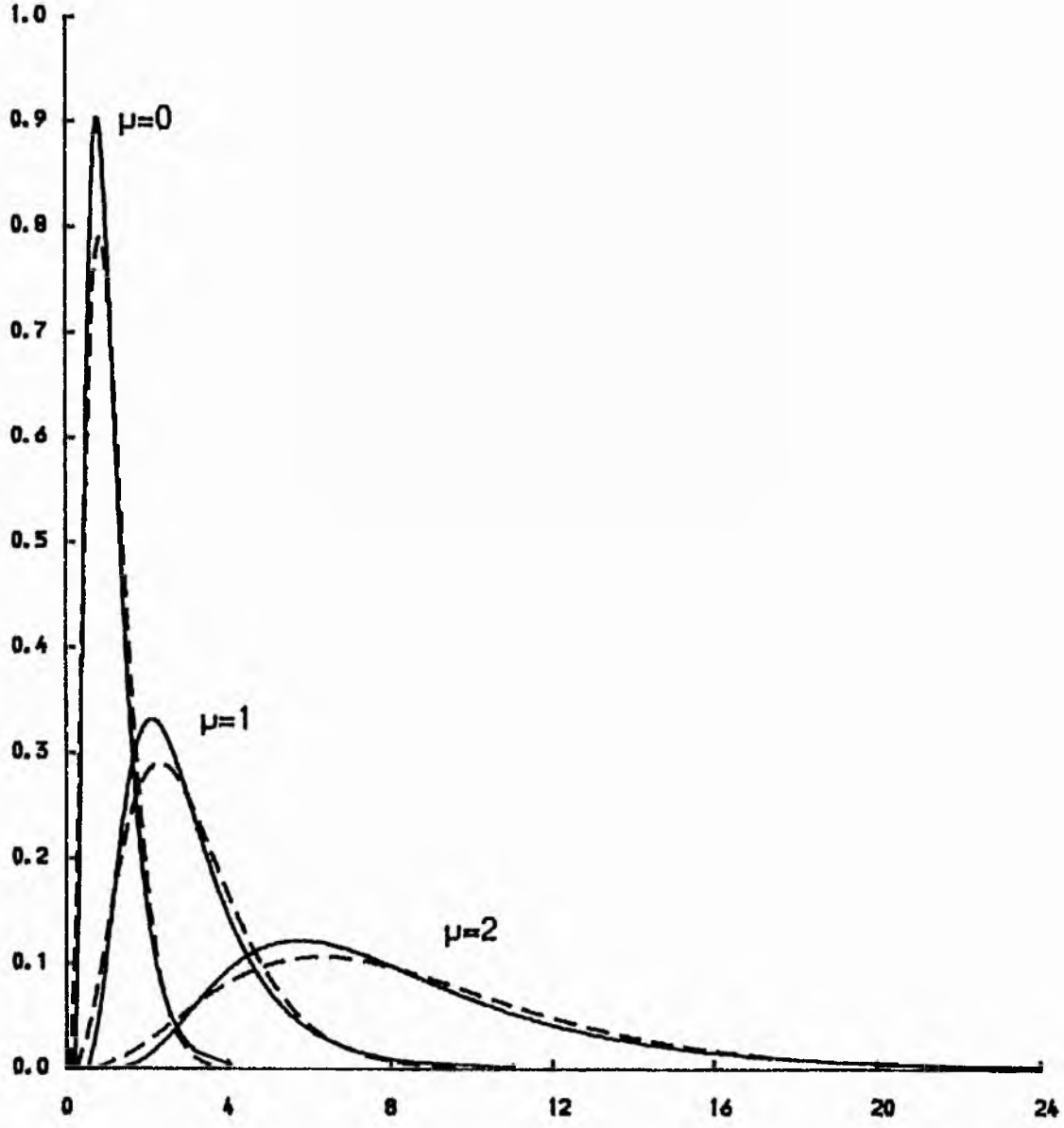


Figure 3.1. Plots of f_{λ^*} (solid lines) and $g_{4,0^*}$ (broken lines) for the Lognormal Distribution ($\sigma=1/2$).

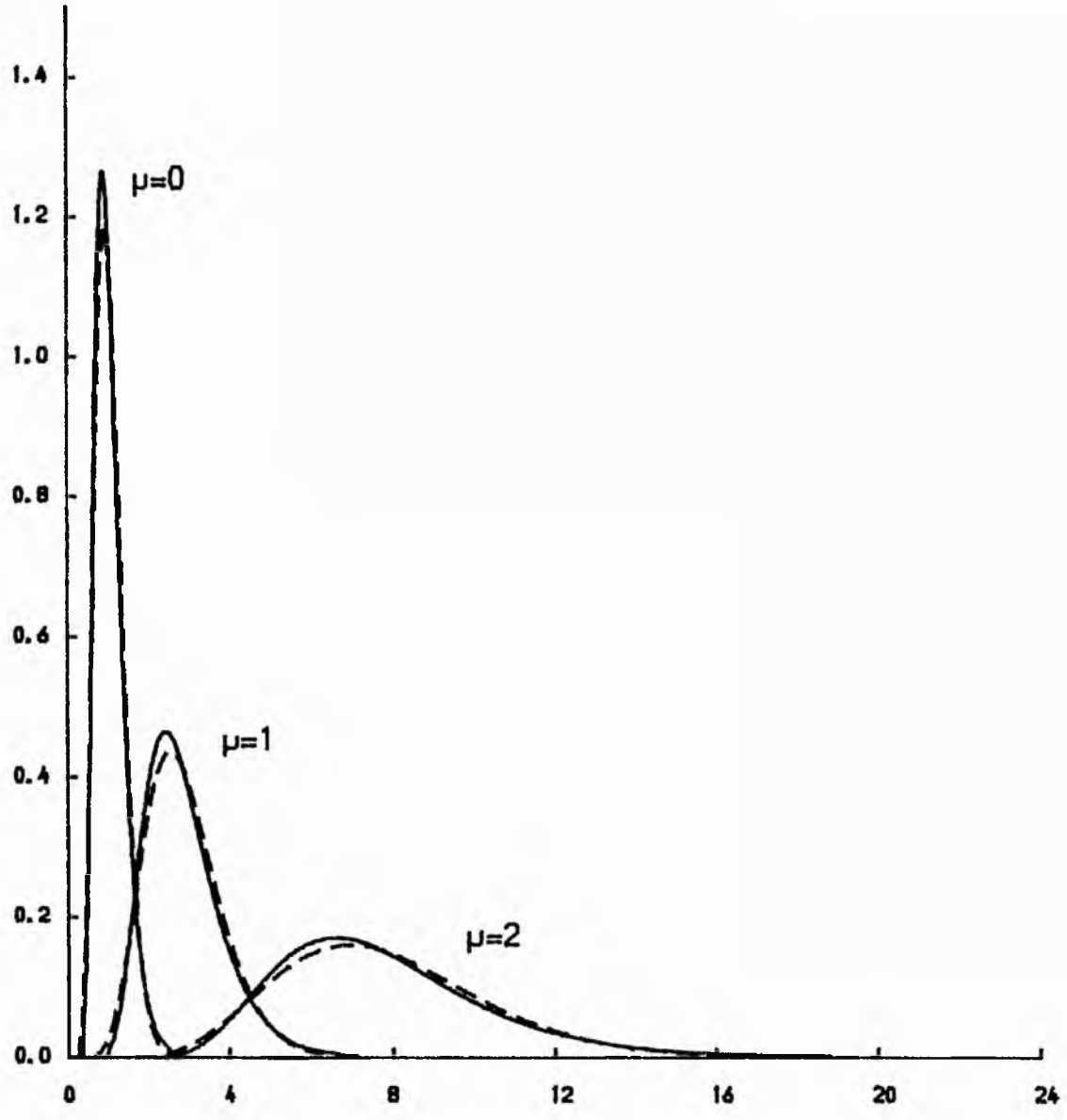


Figure 3.2 Plots of f_{λ^*} (solid lines) and $g_{9,0^*}$ (broken lines) for the Lognormal Distribution ($\sigma=1/3$).

may be used for $0 < k < 1$. This and other asymptotic formulae for the gamma function can be found in Abramowitz and Stegun (1965).

Figure 3.3 shows the 'closeness' of f_{λ^*} to the gamma distribution of order 2 for $\beta=1$ and for $\alpha = 1$, 2 and 4. Figure 3.4 displays f_{λ^*} and g_{4,θ^*} for transforming to the gamma distribution of order 4. The parameter values $\beta=2$ and $\alpha = 1$, 2 and 5 are chosen for inspection.

For the exponential case (m=1)

$$\lambda^* = \beta, \qquad \theta^* = \alpha$$

and

$$K(\lambda^*) = 0.$$

This agrees with the relationship which exists between the Weibull and exponential distributions. If Y is a Weibull (β,α) random variable then Y^{β} has the exponential distribution with scale parameter α.

For the general case, $K(.)$ is a function of β/λ and consequently β/λ^* is constant and equal to k_m say. The power transformation of a Weibull variable is also Weibull and $\lambda^* = 1$ if and only if the Weibull shape parameter β takes the value k_m. Therefore, among all possible Weibull distributions, using the Kullback-Leibler information number as a measure of discrepancy, the Weibull distribution with shape parameter $\beta = k_m$ is 'closest' to the gamma distribution of order m. The gamma scale parameter is $\theta^* = \alpha/m\ \Gamma(1 + 1/k_m)$.

Values of k_m for a range of m values are recorded in Table 3.1. The following expression may be used to approximate k_m for the gamma order m

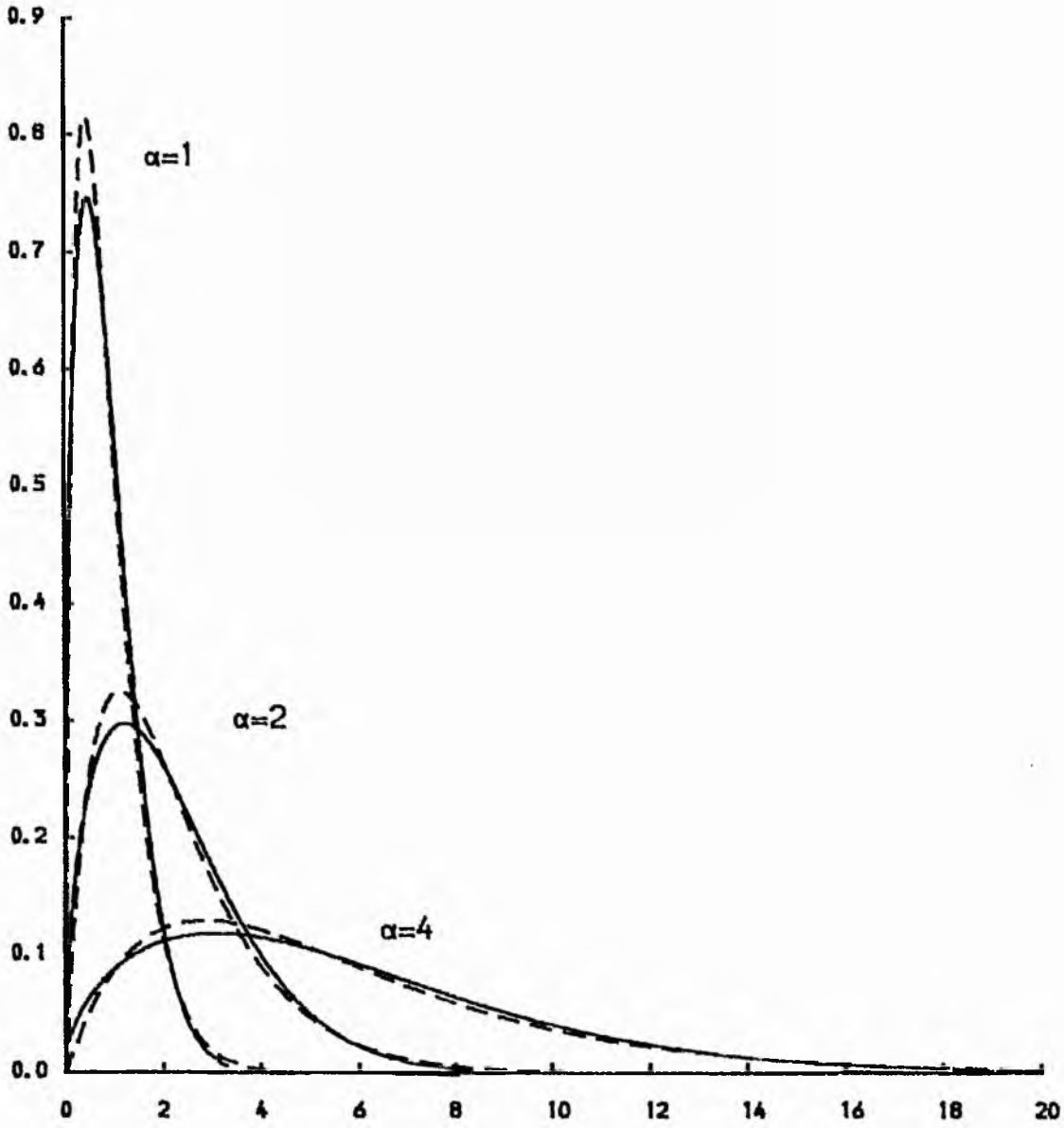


Figure 3.3. Plots of f_{λ^*} (solid lines) and g_{2,θ^*} (broken lines) for the Weibull Distribution ($\beta = 2$).

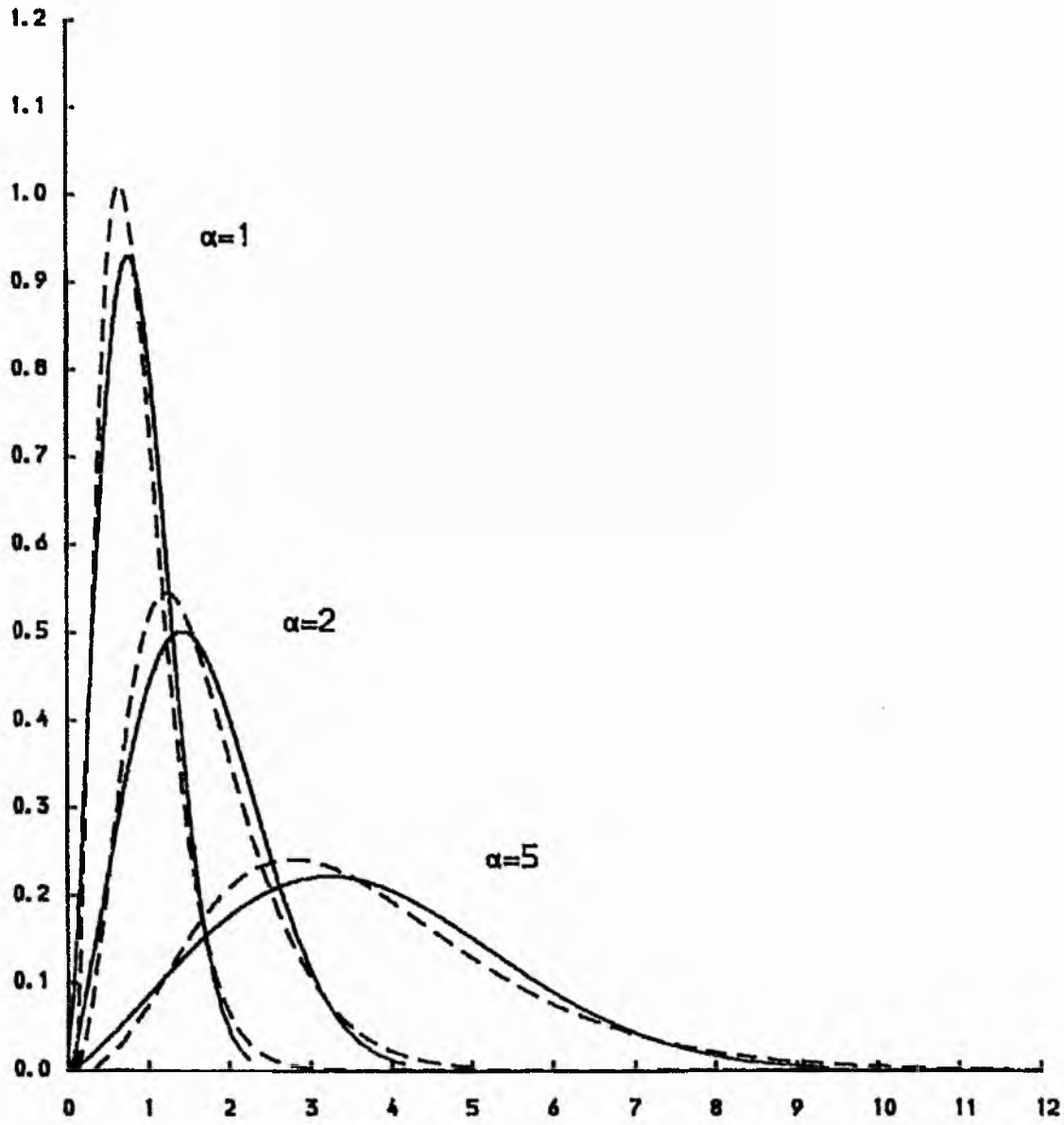


Figure 3.4. Plots of f_{λ^*} (solid lines) and g_{4,θ^*} (broken lines) for the Weibull Distribution ($\beta = 2$).

Table 3.1 Constants k_m for different values of gamma order m

m	k_m
0.2	0.3557
0.4	0.5630
0.6	0.7292
0.8	0.8723
1.0	1.0
2.0	1.5099
3.0	1.9073
4.0	2.2457
5.0	2.5452
6.0	2.8161
7.0	3.0656
8.0	3.2982
9.0	3.5162
10.0	3.7230
12.0	4.1067
14.0	4.4583
16.0	4.7870
18.0	5.0968
20.0	5.3879
25.0	6.0606

$$k_m \approx 0.007087m^{-4} - 0.08002m^{-3} + 0.3328m^{-2} - 0.7081m^{-1}$$
$$+ 1.073 + 0.3915m - 0.01707m^2 + 0.0006016m^3 - 0.00000915m^4. \qquad (3.3.9)$$

This approximation is accurate to three decimal places.

Example 3.3 Pareto family

The Pareto distribution has pdf of the form

$$f(y) = ca^c/y^{c+1}, \quad y \geq a \geq 0, \quad c > 0. \qquad (3.3.10)$$

Since a is a scale parameter, it is sufficient to consider a = 1. Then

$$K(\lambda) = \begin{cases} m(1-\log m) + \log[\Gamma(m)] + m\log[c/(c-\lambda)] + \log(c/\lambda) \\ \qquad - m\lambda/c - 1, \quad \lambda \neq 0 \\ m(1-\log m) + \log[\Gamma(m)] + (1-m)\log c + c\gamma(1-m) - 1, \\ \qquad \lambda = 0. \end{cases} \qquad (3.3.11)$$

The minimising gamma scale parameter as a function of λ is

$$\theta^*(\lambda) = \begin{cases} \frac{1}{m}\, c/(c-\lambda), & \lambda \neq 0 \\ \frac{1}{mc}, & \lambda = 0. \end{cases} \qquad (3.3.12)$$

For transforming to the exponential distribution (m=1)

$$\lambda^* = 0, \quad \theta^* = 1/c$$

and

$$K(\lambda^*) = 0.$$

This agrees with the following relationship which exists between the Pareto and exponential distributions. If Y is a Pareto random

variable then log Y has the exponential distribution with scale parameter 1/c.

For values of m other than one, the value of the transformation parameter which minimises the Kullback-Leibler information number may be readily calculated as

$$\lambda^* = \frac{-1 \pm \sqrt{(1+4m)}}{2m} \, c. \qquad (3.3.13)$$

Transformation estimates and information numbers for various lognormal, Weibull and Pareto distributions are displayed in Table 3.2. The table illustrates numerically, for selected values of m,

Table 3.2 Transforming to the Gamma distribution of known order m : Comparison of transformations and information numbers

m	Distribution	λ^*	$I[f;g_{m,\theta_Y}]$	$I[f_{\lambda^*};g_{m,\theta^*}]$
0.5	Lognormal(μ,0.5)	2.8284	1.0940	0.1532
	Lognormal(μ,2)	0.7071	0.7253	0.1532
1.0	Lognormal(μ,0.5)	2.0	0.6460	0.0809
	Lognormal(μ,2)	0.5	2.0150	0.0809
	Lognormal(μ,1)	0.7071	0.5079	0.0411
	Lognormal(μ,2)	0.3536	5.3690	0.0411
2.0	Weibull(0.5,α)	0.3320	2.6950	0.0061
	Weibull(2,α)	1.3250	0.1865	0.0061
	Pareto(1,2)	1.0	0.6931	0.6931
	Lognormal(μ,1)	0.5	2.2080	0.0206
	Lognormal(μ,2)	0.25	6.1345	0.0206
4.0	Weibull(0.5,α)	0.2230	8.6700	0.0195
	Weibull(2,α)	0.8970	1.6650	0.0195
	Pareto(1,2)	0.7808	0.7123	0.6053
6.0	Lognormal(μ,0.5)	0.8165	0.2563	0.0137
	Lognormal(μ,1)	0.4082	4.4860	0.0137
9.0	Lognormal(μ,0.5)	0.6667	1.0810	0.0090
	Lognormal(μ,1)	0.3333	8.5320	0.0090

the improvement towards the gamma distribution introduced by the transformation for the lognormal and Weibull distributions. Transforming is clearly not successful for the Pareto distribution.

3.4 THE LARGE-SAMPLE BEHAVIOUR OF THE LIKELIHOOD PROCEDURE

Draper and Guttman (1968) investigated transformations of data which allow the transformed data to follow an approximate gamma distribution of known order. They adopted a likelihood approach similar to that of Box and Cox (1964). For the random sample $y_1,\ldots,y_n$, their hypothesis was that for some λ the $z_i = y_i^{\lambda}$, $i = 1,\ldots,n$ follow the gamma distribution of known order m and scale parameter θ to be estimated. For fixed λ the value of θ that maximises the logarithm of the likelihood for the original observations is

$$\hat{\theta}_n(\lambda) = \sum_{i=1}^{n} y_i^{\lambda}/mn. \qquad (3.4.1)$$

Substitution of this maximum likelihood estimate (MLE) for θ in the log-likelihood yields, apart from a constant,

$$L(\lambda) = (m\lambda - 1) \sum_{i=1}^{n} \log y_i - mn \log \left[\sum_{i=1}^{n} y_i^{\lambda}/n \right] + n \log \lambda. \qquad (3.4.2)$$

The function $L(.)$ can be plotted against λ and the maximising value $\hat{\lambda}_n$ estimated from the plot.

If $y_1,\ldots,y_n$ is a random sample from $f(.)$, then for large samples and general function $h(.)$, $\frac{1}{n}\sum_{i=1}^{n} h(y_i) \approx E_f[h(y)]$. This implies that for large n and suitable choices of h, the following approximations hold for the MLE of θ and the log-likelihood $L(\lambda)$

$$\hat{\theta}_n(\lambda) \approx \frac{1}{m} E_f(y^{\lambda}), \qquad (3.4.3)$$

$$L(\lambda) \approx n(m\lambda - 1)E_f(\log y) - mn \log[E_f(y^{\lambda})] + n \log \lambda. \qquad (3.4.4)$$

But these large-sample approximations of $\hat{\theta}_n(\lambda)$ and $L(\lambda)$ are $\theta^*(\lambda)$ and $-nK(\lambda)$ (apart from a constant) respectively. This means that the maximum likelihood estimates, $\hat{\theta}_n(\hat{\lambda}_n)$ and $\hat{\lambda}_n$, converge to the values, $\theta^*(\lambda^*)$ and λ^*, that minimise the Kullback-Leibler information number between $f_{\lambda}(.)$, the true density for the transformed variable and a gamma distribution of known order.

A simulation study is conducted to illustrate this result for the transformation parameter. Two hundred random samples are generated from selected distributions for each of the sample sizes 20, 40, 60, 80, 100 and 150. For each sample the MLE of λ for transforming to a gamma distribution of known order m, is calculated. The mean values (together with the standard deviations) are recorded in Table 3.3 for comparison with the estimates of λ obtained on minimising the information number. Table 3.3 shows clearly that the value of λ that minimises the information number is the limiting value of the MLE.

Table 3.3 Transforming to the Gamma distribution of known order m:

Mean value (and standard deviation) for MLE of λ

m	Distribution	n = 20	40	60	80	100	150	λ^*
1.0	Lognormal(0,2)	0.562 (0.108)	0.534 (0.077)	0.525 (0.067)	0.518 (0.055)	0.514 (0.051)	0.509 (0.040)	0.5
2.0	Lognormal(0,1)	0.756 (0.146)	0.733 (0.103)	0.729 (0.083)	0.726 (0.067)	0.719 (0.065)	0.716 (0.048)	0.707
2.0	Weibull(2,1)	1.381 (0.196)	1.362 (0.142)	1.347 (0.125)	1.345 (0.098)	1.337 (0.093)	1.324 (0.079)	1.325
4.0	Lognormal(0,0.3)	1.748 (0.203)	1.727 (0.170)	1.719 (0.150)	1.704 (0.107)	1.677 (0.101)	1.675 (0.090)	1.677
4.0	Weibull(0.5,1)	0.237 (0.051)	0.229 (0.029)	0.225 (0.022)	0.225 (0.022)	0.224 (0.020)	0.224 (0.016)	0.223

3.5 CONCLUDING REMARKS

This chapter presents an information approach for transforming variables with known distributions to follow the gamma distribution of known order, with applications including the lognormal, Weibull and Pareto distributions. A transformation is determined by considering a parametric family of power transformations and employing the Kullback-Leibler information number as a measure of discrepancy between two probability distributions. The transformation parameter and the gamma scale parameter are chosen to minimise the information number between the true density of the transformed variable and the gamma distribution of known order. The parameter values obtained are the limiting values of the maximum likelihood estimates resulting from the likelihood procedure of Draper and Guttman (1968) for transforming data to follow a gamma distribution of known order. A simulation study was undertaken to illustrate the large-sample behaviour of the maximum likelihood estimator and the results are presented. Finally, the information number approach allows us to determine the lognormal and Weibull distributions that are hardest to discriminate from a gamma distribution of known order.

CHAPTER 4

THE RATIO PROCEDURE FOR MODEL TESTING AND ESTIMATION

4.1 INTRODUCTION

Graphical techniques are powerful tools for investigating data and displaying information on the underlying distribution. They are useful for determining the goodness-of-fit of the data to a hypothesised distribution and for detecting outliers or contamination. Some techniques may also be used to supply parameter estimates within a model.

The purpose of this chapter is to present a new graphical procedure. The procedure applies to any distribution (with associated random variable Y) for which the following holds: some power transformation of Y has a distribution which does not include an unknown shape parameter. Therefore, application of the procedure is feasible for important parametric models including the exponential, uniform, Pareto, Weibull, Gumbel, normal, lognormal, logistic and log-logistic distributions. If the hypothesised distribution does not include an unknown shape parameter the new technique is more informative than other established techniques, in that when the hypothesised distribution is rejected, the graphs provide an estimate of the transformation to that distribution.

The technique is based on a generalisation of a ratio R(y) introduced by Tarter and Kowalski (1972) who defined R(y) to test for and suggest transformations to normality. Their definition is

$$R(y) = \frac{\phi\Phi^{-1}F(y)}{f(y)} , \qquad (4.1.1)$$

where f and ϕ represent the density function of a given random sample and the standard normal density respectively, with corresponding cumulative distribution functions F and Φ.

Tarter and Kowalski used the following results to relate the ratio R(y) to transformations to normality.

Let (a,b) be an open interval on which f(y) and $\phi(y)$ are continuous and non-zero. Then

(1) If there exists a function $\psi(y)$ such that $\psi'(y) = d\psi(y)/dy$ is continuous and non-zero on (a,b) and if $F(y) = \Phi[\psi(y)+\kappa]$ on (a,b) for fixed constant κ, then $R(y) = 1/\psi'(y)$ on (a,b).

Conversely

(2) If $R(y) = 1/\psi'(y)$ is continuous and non-zero on (a,b) then $F(y) = \Phi[\psi(y)+\kappa]$ on (a,b) for fixed constant κ.

Tarter and Kowalski dealt with the problem of estimating R(y) by modifying the procedure for the estimation of $1/fF^{-1}(t)$ which was considered by Siddiqui (1960) and Bloch and Gastwirth (1968). For the ordered sample $y_1<y_2<\ldots<y_n$ from density f, Tarter and Kowalski estimated the graph of R(y) by plotting $R_i = (y_{i+1}-y_i)\phi\Phi^{-1}(p_i)/\delta$ against $(y_i+y_{i+1})/2$ for i=1,2,...,n-1, where

$p_i = i/n$ is the plotting position and $\delta = 1/n$ is the distance between two adjacent plotting positions. The argument upon which the method is based is described in Appendix C. Tarter and Kowalski suggested that the original procedures of Siddiqui (1960) and Bloch and Gastwirth (1968) should be used to graph R(y) when the sample size is moderate to large, i.e. n>40.

4.2 GENERALIZATIONS

The definition of the ratio R(y) may be generalised to provide a test for any hypothesised distribution (with associated variable Y) for which the following holds: some power transformation of Y has a distribution whose cumulative G say, does not include an unknown shape parameter. For the hypothesised distribution the generalisation is of the form

$$R(y) = \frac{gG^{-1}F(y)}{f(y)} , \qquad (4.2.1)$$

where G is standardised and g is the corresponding standard probability density function. The properties of the ratio R(y) for the hypothesised distribution provide a means of checking the appropriateness of the model.

When G is the uniform cumulative on the interval [0,1]

$$R(y) = 1/f(y) \qquad (4.2.2)$$

and the ratio is simply the reciprocal of the probability density function of the random sample. By result (2) of Section 4.1, with the normal distribution replaced by the uniform distribution, the transformation to the uniform distribution is

$$\psi(y) = F(y), \qquad (4.2.3)$$

the well-known Probability Integral Transformation. This provides a ready proof for the Probability Integral Transformation theorem [Bury (1975)].

When G is the standard exponential cumulative the ratio reduces to

$$R(y) = \frac{1-F(y)}{f(y)}, \qquad (4.2.4)$$

which is called Mill's Ratio in economics and is simply the reciprocal of the failure rate or hazard function. The hazard function is a useful concept in reliability studies, studies of mortality and seismology. A plot of the hazard function is often preferable to plotting the log survivor function as the hazard function has the advantage that it often varies slowly over all or most of the range. This can prove useful in selecting a distribution model. The ratio plot may be assessed as the inverted plot of the hazard function.

The ratio R(y) as expressed in (4.2.1) is the reciprocal of the generalised failure rate function as defined by Barlow and Van Zwet (1969).

When the hypothesised distribution does not include an unknown shape parameter and graphically testing with the ratio R(y) leads to rejection, R(y) may be used to suggest transformations to that distribution. Tarter and Kowalski relied on visual inspection of the estimated graphs of R(y) to suggest transformations to normality. Therefore, the ratio-concept may be further

generalised by considering a family of power transformations

$$\Psi(y) = \begin{cases} y^{\lambda}, & \lambda \neq 0 \\ \log y, & \lambda = 0. \end{cases} \qquad (4.2.5)$$

Plotting log R(y) against log y, called a log-ratio plot, stabilises the variance and is linear with slope $1-\lambda$. The slope of the straight line fitted to the plot may be used to obtain an estimate of the transformation parameter λ, which in turn may be used to transform to the hypothesised distribution.

The ratio plot is estimated by plotting

$R_i = (y_{i+1}-y_i)gG^{-1}(p_i)/\delta$ against $(y_i+y_{i+1})/2$ for $i=1,2,\ldots,n-1$,

where p_i is the plotting position and δ is the distance between two adjacent plotting positions. Sutcliffe et al. (1975) recommended the plotting position $p_i = (i-0.44)/(n+0.12)$ [Gringorten (1963)] for general use with the exponential and Gumbel (extreme value Type 1) distributions and the plotting position $p_i = (i-\frac{3}{8})/(n+\frac{1}{4})$ [Blom (1958)] for general use with the normal and lognormal distributions. In the sections that follow applications for the ratio procedure to data will use these recommended plotting positions.

The log-ratio plot is formed by plotting the n-1 points

$$(\log[(y_i+y_{i+1})/2],\ \log R_i),$$

$i=1,2,\ldots,n-1$.

Ratio plotting extends naturally to singly Type 1 and Type 2 censored data. The plotting positions of the observations are the same as when the complete sample is available. Points corresponding to censored observations are not plotted. Otherwise the procedure applies directly.

4.3 TRANSFORMING TO THE EXPONENTIAL DISTRIBUTION

As seen in Section 4.2, an interesting choice for G is the standard exponential cumulative

$$G(y) = 1-\exp\{-y\}. \qquad (4.3.1)$$

The practical contribution of the ratio R(y) lies in the following properties which are immediate consequences of results (1) and (2) of Section 4.1, adapted for transforming to the exponential distribution.

Property 1. $R(y) = \alpha$ for all finite y if and only if F(y) is the exponential cumulative with scale parameter α.

For property 1, the ratio plot shows a horizontal band of points scattered about the value of the exponential scale parameter. If the exponential distribution is rejected the transformation parameter λ may be estimated by fitting a straight line to the log-ratio plot and calculating the slope. The estimate may then be used to transform to the exponential distribution.

Property 2. $R(y) = \tau-y$ for all finite y if and only if $F(y) = G[-\log(1-y/\tau)]$ i.e. if and only if f(y) is the uniform probability density function on the interval $[0,\tau]$.

For property 2, the ratio plot is linear with slope -1 and provides an estimate of the parameter τ when the uniform distribution appears reasonable. This is obtained by fitting a straight line of slope -1 to the plot and estimating τ as the intercept on the ordinate axis.

Property 3. $R(y) = y/c$ for all finite y if and only if $F(y) = G[\log y]$ i.e. if and only if F(y) is the Pareto cumulative with probability density function

$$f(y) = c/y^{c+1} . \qquad (4.3.2)$$

For property 3, the ratio plot is linear and passes through the origin. The Pareto parameter c may be estimated using the slope of the fitted line. In addition, the log-ratio plot is linear with slope 1, which is indicative of the log transformation.

Property 4. $R(y) = 1/\beta\ \alpha^{\beta}\ y^{1-\beta}$ for all finite y if and only if $F(y) = G[(y/\alpha)^{\beta}]$ i.e. if and only if F(y) is the two-parameter Weibull distribution with probability density function

$$f(y) = \beta/\alpha\,(y/\alpha)^{\beta-1}\exp\{-(y/\alpha)^{\beta}\}. \qquad (4.3.3)$$

For property 4, the ratio plot shows a power function in y, monotone and increasing in y if $\beta<1$ and monotone decreasing if $\beta>1$. The log-ratio plot is linear with slope $1-\beta$ where β is the Weibull shape parameter. The intercept of the fitted line on the ordinate axis may be used to calculate an estimate of the Weibull scale parameter α.

An example illustrating the ratio procedure for the exponential case concludes this section.

Example 4.1 CVDSC Data: The data is from Jones and Rowcliffe (1979). A sample of 30 tensile strengths for chemical vapor-deposited silicon carbide (CVDSC) were obtained from expanded-ring tests. The tensile strengths were

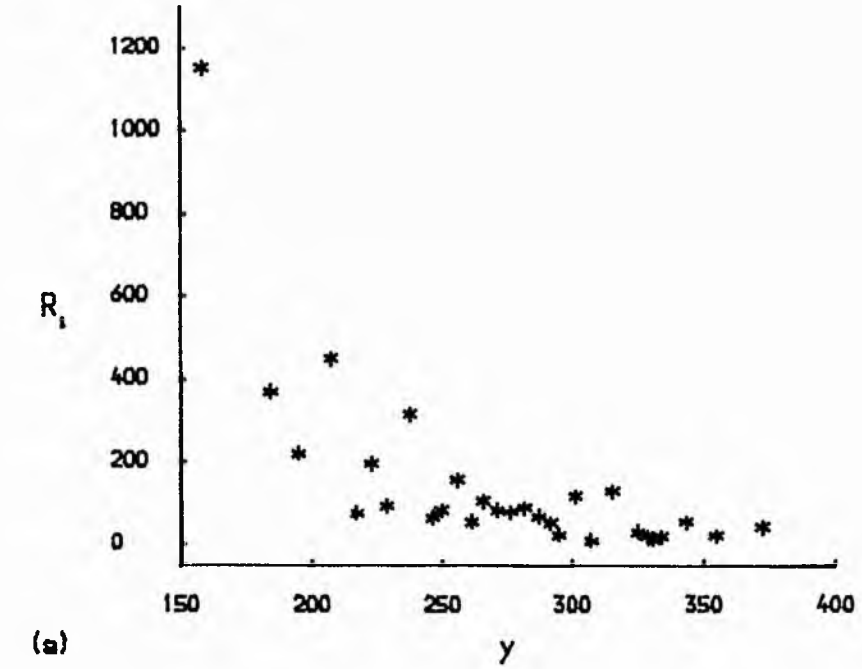
386	359	351	336	332	330	327	323
308	307	296	294	290	285	279	274
269	263	260	252	248	245	231	227
219	216	199	191	178	139		

The Weibull distribution has often been found appropriate in such a situation. The cumulative failure probability plot of Jones and Rowcliffe yielded estimates taking the values 285.0 and 5.4 for the two-parameter Weibull scale and shape parameters respectively. The corresponding maximum likelihood estimates are 296.6 and 5.6.

The ratio plot is estimated as outlined in Section 4.2 (with G the standard exponential cumulative). The plot, shown in Figure 4.1a, is monotone and decreasing which suggests that a two-parameter Weibull distribution with shape parameter greater than one is reasonable. The plot also reveals that the smallest observation is most influential.

The straight-line log-ratio plot of Figure 4.1b confirms that the data are drawn from a two-parameter Weibull distribution. For the Weibull distribution $\log R(y) = -\log \beta + \beta \log \alpha + (1-\beta) \log y$. If a straight line is drawn through the plot, β can be estimated from the slope and α estimated from the intercept. A line is fitted to the plot using unweighted least squares regression (alternatively any other formal method). The slope of the line is approximately 4.4 giving $\beta_R = 5.4$ as an estimate of the shape parameter (subscript R represents ratio estimation). The intercept on the ordinate axis is 28.84 giving $\alpha_R = 290.4$.

The ratio estimates obtained here are quite close to the estimates of Jones and Rowcliffe. This is no surprise as the cumulative failure probability plot is based on the failure rate



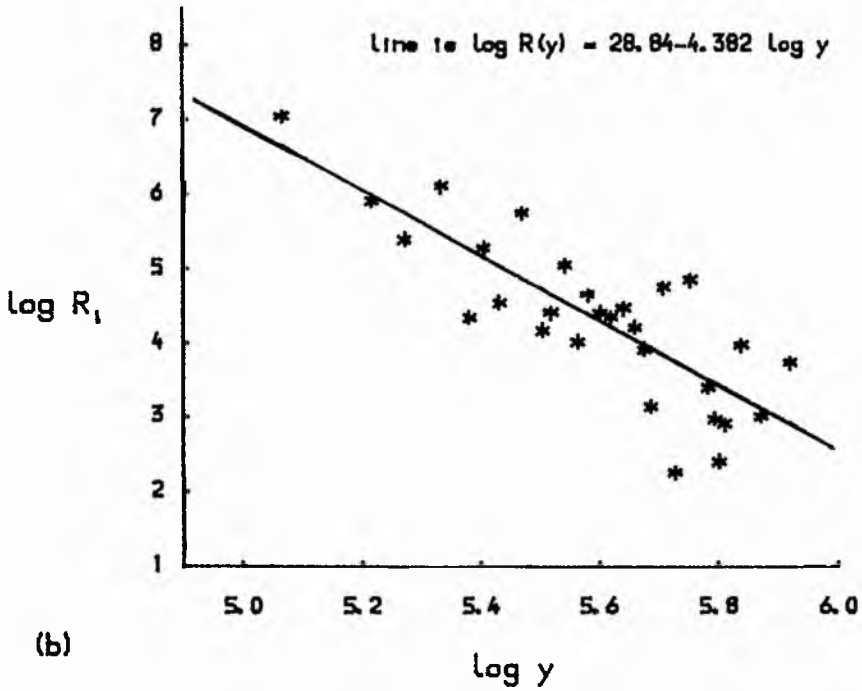


Figure 4.1. Exponential analysis of CVDSC Data

(a) Estimated ratio plot

(b) Estimated log-ratio plot

or hazard function and the ratio plot for transforming to the exponential distribution is simply the inverted plot of the hazard function.

4.4 TRANSFORMING TO THE GUMBEL AND NORMAL DISTRIBUTIONS

Another interesting choice for G is the standard Gumbel cumulative

$$G(y) = 1-\exp\{-\exp y\}. \qquad (4.4.1)$$

The two-parameter Weibull distribution as defined in the last section, is related to the Gumbel distribution as follows: log Y has the Gumbel distribution with location parameter $\nu = \log \alpha$ and scale parameter $\theta = 1/\beta$.

The following properties describe the practical value of the ratio R(y) for the Gumbel analysis.

Property 1. $R(y) = \theta$ for all finite y if and only if F(y) is the Gumbel cumulative with scale parameter θ.

For property 1, the ratio plot shows a horizontal band of points scattered about the Gumbel scale parameter value. If the Gumbel distribution is rejected, the transformation parameter λ may be estimated from the slope of the line fitted to the log-ratio plot and the estimate used to transform to the Gumbel distribution.

Property 2. $R(y) = y/\beta$ if and only if F(y) is the two-parameter Weibull cumulative i.e. if and only if $F(y) = G[\log((y/\alpha)^{\beta})]$.

For property 2, the ratio plot is linear, which is indicative of the log transformation, and passes through the origin. When the two-parameter Weibull distribution is appropriate, the slope of the line fitted to pass through the origin provides an estimate of $1/\beta$. In addition, the log-ratio plot is linear with slope 1 ($=1-\lambda$). Least squares regression may be used to fit lines to plots.

A non-zero intercept for the ratio plot indicates that there is a non-zero threshold value γ. The log-ratio plot will not produce a straight line of slope one unless the γ value or an estimate of it is subtracted from the data and the log of the shifted data is treated as the abscissa. This suggests the following iterative procedure, based on interpolation, for estimating the threshold parameter:

Let λ_j be the estimate of the transformation parameter obtained from the slope of the log-ratio plot, where the abscissa is the log of the data shifted to the left by the amount γ_j. With starting values for the threshold of $\gamma_0 = 0$ and $\gamma_1 =$ smallest sample value, the following expression may be used iteratively to give an estimate of the threshold parameter

$$\gamma_{j+1} = \begin{cases} \gamma_j - \lambda_j(\gamma_j - \gamma_{j-1})/(\lambda_j - \lambda_{j-1}), & \gamma_{j+1} \leq Y_1 \\ x_1 & , \gamma_{j+1} > Y_1 \end{cases} \qquad (4.4.2)$$

for $j \geq 1$.

The ratio estimate γ_R emerges after a few iterations. The data minus γ_R may now be treated as having a two-parameter Weibull distribution and the ratio estimate β_R calculated from the slope

of the line fitted to the ratio plot for the shifted data.

The ratio estimates of the threshold and shape parameters of the three-parameter Weibull distribution are calculated independently of the scale parameter. As yet no ratio estimate for the scale parameter has been proposed. The maximum likelihood estimate (MLE) of the scale parameter is

$$\hat{\alpha} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{\gamma})^{1/\hat{\beta}}, \qquad (4.4.3)$$

where $\hat{\gamma}$ and $\hat{\beta}$ are the MLE's of the threshold and shape parameters respectively.

A suggested ratio estimate for α is given by (4.4.3) with the MLE's of γ and β replaced by the ratio estimates.

The properties of the ratio for the Weibull distribution as described above hold identically for the original ratio of Tarter and Kowalski (1972) in the case of the three-parameter lognormal distribution with probability density function of the form

$$f(y) = [\beta(2\pi)^{-\frac{1}{2}}(y-\gamma)^{-1}] \exp\{-\beta^2(\log[(y-\gamma)/\alpha])^2/2\}. \qquad (4.4.4)$$

The random variable log Y has a normal distribution with mean $\mu = \log \alpha$ and standard deviation $\sigma = 1/\beta$. The ratio estimates γ_R and β_R are obtained using the procedure already outlined for the Weibull distribution. The suggested ratio estimate for the scale parameter is

$$\alpha_R = \exp\{\frac{1}{n} \sum_{i=1}^{n} \log(y_i - \gamma_R)\}. \qquad (4.4.5)$$

This is the MLE of scale with the MLE of threshold replaced by the ratio estimate.

4.5 A SIMULATION STUDY

The ratio estimation procedure for the Weibull distribution is evaluated and compared with the modified maximum likelihood estimation procedure of Cohen and Whitten (1982) by means of a simulation study. Cohen and Whitten solved three nonlinear equations simultaneously to obtain their modified maximum likelihood estimates (MMLE's). Two of these equations are found by equating to zero the partial derivatives of the likelihood function with respect to α and β. The third equation is found by equating $F(y_1)$ to its expected value, $1/(n+1)$ where F is the Weibull cumulative distribution function and y_1 is the first order statistic. An optimisation replaces the "trial and error" procedure which Cohen and Whitten used together with linear interpolation to solve the estimating equations. The MMLE procedure sometimes fails to find solutions to the equations.

The ratio estimators have two important advantages over MMLE's. They do not involve the numerical solution of nonlinear equations and estimates always emerge.

The simulation study was set up as follows. The paired scale and shape parameter values $(\alpha,\beta) = (2.0,0.8)$, $(5.0,1.2)$ and $(10.0,2.0)$ were chosen for investigation. For each (α,β) value two hundred random samples consisting of 20 observations each were generated from a Weibull distribution with zero threshold and scale and shape parameters equal to the (α,β) value. In addition, two hundred random samples each of size 20 were generated from a Weibull distribution with threshold parameter equal to 5.0 and

scale and shape parameters equal to the (α,β) value, for each (α,β) value. The entire process was repeated for a sample size of 40.

For each sample the MMLE's and the ratio estimates for the threshold and shape parameters were obtained. In calculating estimates all three parameters were assumed to be unknown. The bias and root mean squared error estimates are entered in Tables 4.1 and 4.2. For each estimator, results recorded are based only on samples which yielded estimates. In each three-line entry the top line pertains to the threshold parameter, the second to the shape parameter and the third to the number of samples which failed to yield estimates.

For the cases where β = 0.8 and 1.2 there is not much to choose between the ratio estimators and the MMLE's. Where the MMLE biases and root mean squared errors are smaller than their ratio counterparts, it should be pointed out that samples had been discarded for the MMLE procedure. For the case where β = 2.0 the biases and root mean squared errors for the ratio γ and β estimators are considerably smaller than those for the MMLE γ and β estimators.

Kappenman (1985) illustrated his closed form procedure for estimating the parameters of the Weibull and lognormal distributions by generating a random sample of size 25 from each of the two distributions. These samples provide a means of checking the calculation of ratio estimates. For each sample the threshold and scale parameters were set equal to zero and one respectively. The shape parameter $c(=\beta)$ for the Weibull

Table 4.1 Simulation Summary - Bias for parameter estimators

Estimator	200 samples n=20						200 samples n=40					
	γ=0.0 α=2.0 β=0.8	γ=5.0 α=2.0 β=0.8	γ=0.0 α=5.0 β=1.2	γ=5.0 α=5.0 β=1.2	γ= 0.0 α=10.0 β= 2.0	γ= 5.0 α=10.0 β= 2.0	γ=0.0 α=2.0 β=0.8	γ=5.0 α=2.0 β=0.8	γ=0.0 α=5.0 β=1.2	γ=5.0 α=5.0 β=1.2	γ= 0.0 α=10.0 β= 2.0	γ= 5.0 α=10.0 β= 2.0
MMLE	0.014 -0.060 17	0.008 -0.070 13	-0.081 0.038 12	-0.109 0.029 6	-1.877 0.473 8	-1.180 0.329 4	0.026 -0.020 3	0.005 -0.018 2	-0.006 -0.038 1	-0.011 -0.001 0	-0.654 0.158 2	-0.645 0.173 0
Ratio	0.010 0.011 0	0.030 -0.040 0	0.139 0.013 0	0.183 -0.024 0	0.190 0.044 0	0.332 0.020 0	0.015 -0.061 0	0.009 -0.009 0	0.08 -0.066 0	0.108 -0.049 0	0.411 -0.110 0	0.381 -0.092 0

NOTE: In each three-line entry the top line pertains to the threshold parameter γ, the second to the shape parameter β and the third to the number of samples which failed to yield estimates.

Table 4.2 Simulation Summary - Root mean squared error for parameter estimators

	200 samples n=20						200 samples n=40					
Estimator	γ=0.0 α=2.0 β=0.8	γ=5.0 α=2.0 β=0.8	γ=0.0 α=5.0 β=1.2	γ=5.0 α=5.0 β=1.2	γ= 0.0 α=10.0 β= 2.0	γ= 5.0 α=10.0 β= 2.0	γ=0.0 α=2.0 β=0.8	γ=5.0 α=2.0 β=0.8	γ=0.0 α=5.0 β=1.2	γ=5.0 α=5.0 β=1.2	γ= 0.0 α=10.0 β= 2.0	γ= 5.0 α=10.0 β= 2.0
MMLE	0.073 0.193 17	0.083 0.220 13	0.841 0.500 12	0.621 0.399 6	4.955 1.537 8	4.371 1.399 4	0.133 0.186 3	0.038 0.123 2	0.205 0.193 1	0.270 0.218 0	2.529 0.876 2	1.921 0.641 0
Ratio	0.231 0.366 0	0.080 0.275 0	0.873 0.530 0	0.545 0.385 0	3.406 1.144 0	4.604 1.504 0	0.122 0.232 0	0.049 0.235 0	0.248 0.284 0	0.283 0.270 0	1.969 0.694 0	1.705 0.548 0

NOTE: In each three-line entry the top line pertains to the threshold parameter γ, the second to the shape parameter β and the third to the number of samples which failed to yield estimates.

distribution was set equal to 1.5 and the lognormal shape parameter $c(=1/\beta)$ was chosen to be 0.8. The generated samples are presented in Table 4.3.

Kappenman recorded his estimates for the sampled Weibull and lognormal distributions in a table for comparison with the true values (TV's) and the modified maximum likelihood estimates (MMLE's) of Cohen and Whitten (1982). Also included are the weighted least squares estimates (WLSE's) proposed by Munro and Wixley (1970) for the lognormal sample. Estimates for the parameters of the Weibull and lognormal samples are found using the ratio procedure. All sets of estimates are recorded in Table 4.4. It can be seen that, with the exception of the Weibull scale parameter, the ratio estimates obtained are quite good.

The next section presents two examples in which the ratio graphical procedure is applied to real data.

4.6 APPLICATIONS

Example 4.2 Grubbs Data: Grubbs (1971) gives mileages at which nineteen military personnel carriers failed. The mileages were 162, 200, 271, 302, 393, 508, 539, 629, 706, 777, 884, 1008, 1101, 1182, 1463, 1603, 1984, 2355, 2880. The two-parameter exponential model is believed to be suitable for this data. Lawless (1977) used this data set to illustrate how to obtain prediction intervals for the two-parameter exponential distribution.

The ratio plot for the Gumbel analysis is shown in Figure

Table 4.3 Random samples of size 25 from the Weibull and lognormal distributions with $\gamma=0$ and $\alpha=1$

	0.030	0.150	0.229	0.242	0.269	0.315	0.318	0.411
Weibull	0.425	0.479	0.528	0.567	0.598	0.690	0.800	0.883
c=1.5	0.947	0.982	0.994	1.067	1.158	1.465	1.514	1.656
	1.978							
	0.284	0.327	0.412	0.555	0.619	0.680	0.703	0.729
Lognormal	0.801	0.838	0.843	0.943	1.024	1.026	1.099	1.613
c=0.8	1.656	1.678	1.751	1.771	2.166	2.426	2.523	2.841
	4.056							

Table 4.4 Summary of estimates

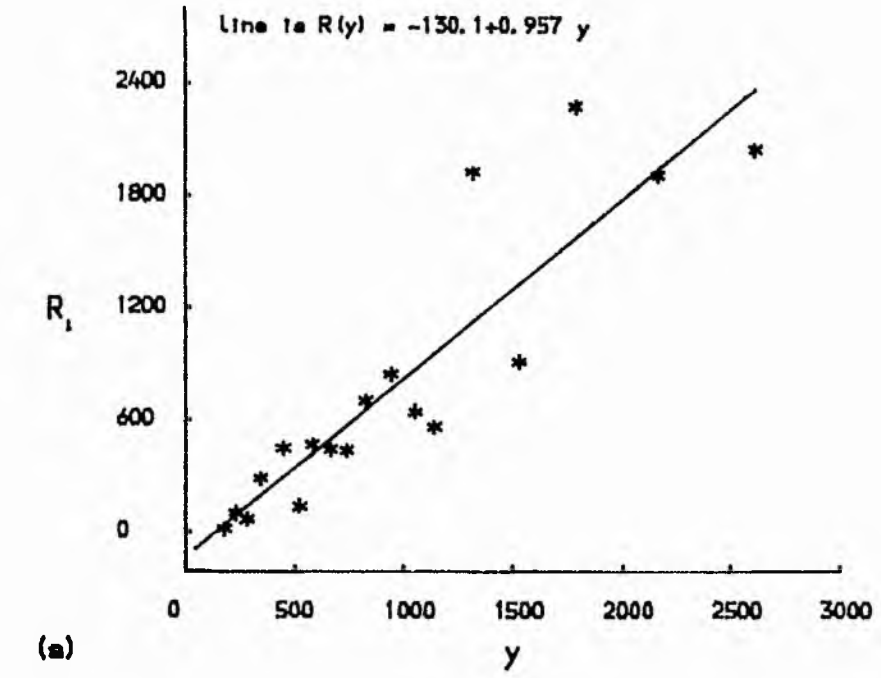
		γ	α	c
Weibull	TV	0	1	1.5
	MMLE	−0.121	0.982	1.833
	Kappenman	−0.080	0.962	1.609
	Ratio	0.029	0.782	1.375
Lognormal	TV	0	1	0.8
	MMLE	−0.016	1.094	0.659
	WLSE	−0.033	1.119	0.723
	Kappenman	0.108	0.934	0.772
	Ratio	0.042	1.021	0.775

4.2a. The plot is roughly linear suggesting that the Weibull distribution is reasonable. The plot does not pass through the origin and the line fitted to the log-ratio plot of Figure 4.2b has slope 1.511 giving $\lambda_R = -0.511$ as an estimate of the transformation parameter. Therefore, the evidence is for a non-zero threshold value. The log-ratio plot clearly shows that the smallest observation is heavily influencing that evidence. The ratio estimate of the threshold parameter is calculated using (4.4.2) to be $\gamma_R = 155.3$.

The log-ratio plot for the data minus this ratio estimate, as shown in Figure 4.3b, is linear with slope one and no departures from the straight line. The ratio plot for the shifted data is shown in Figure 4.3a. The fitted straight line has slope $1/\beta_R = 0.970$ which gives $\beta_R = 1.031$. The ratio estimate of the scale parameter is calculated using the expression (4.4.3) with the MLE's of threshold and shape replaced by the ratio estimates. The estimate takes the value 851.7.

By comparison, the MMLE's of Cohen and Whitten (1982) are found to be $\hat{\gamma} = 113.2$, $\hat{\beta} = 1.038$ and $\hat{\alpha} = 853.7$.

The goodness-of-fit for the two Weibull distributions, specified by the two sets of parameter estimates, is tested using the empirical distribution function statistics, the Cramer-von Mises statistic W^2 and the Anderson-Darling statistic A^2 [Stephens (1977)]. When the parameters are estimated using the ratio procedure the values of the statistics are $W^2 = 0.014$ and $A^2 = 0.148$. For modified maximum likelihood estimation, the values of the test statistics are $W^2 = 0.029$ and $A^2 = 0.180$.



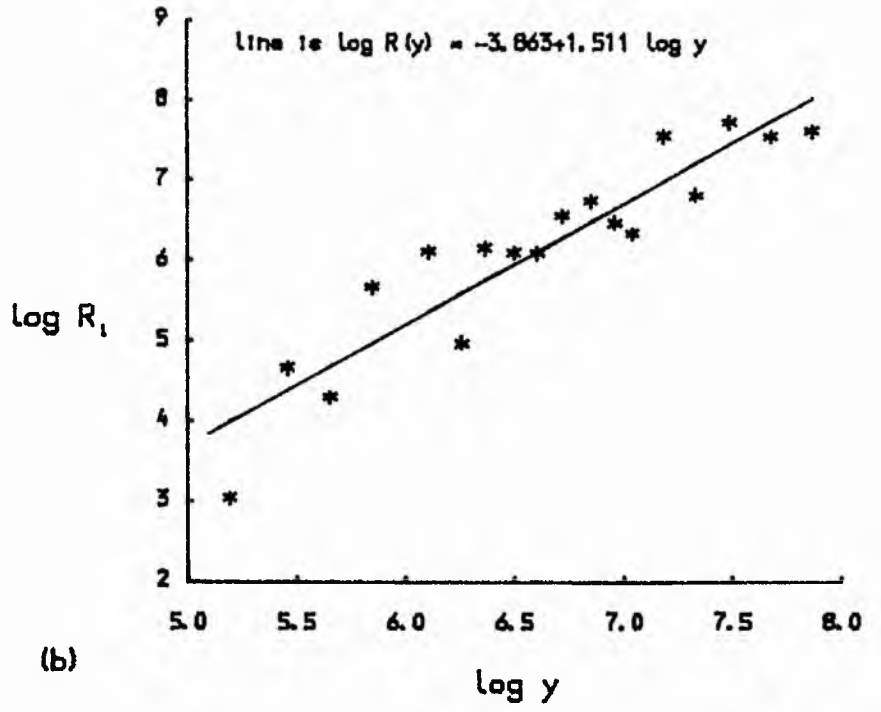
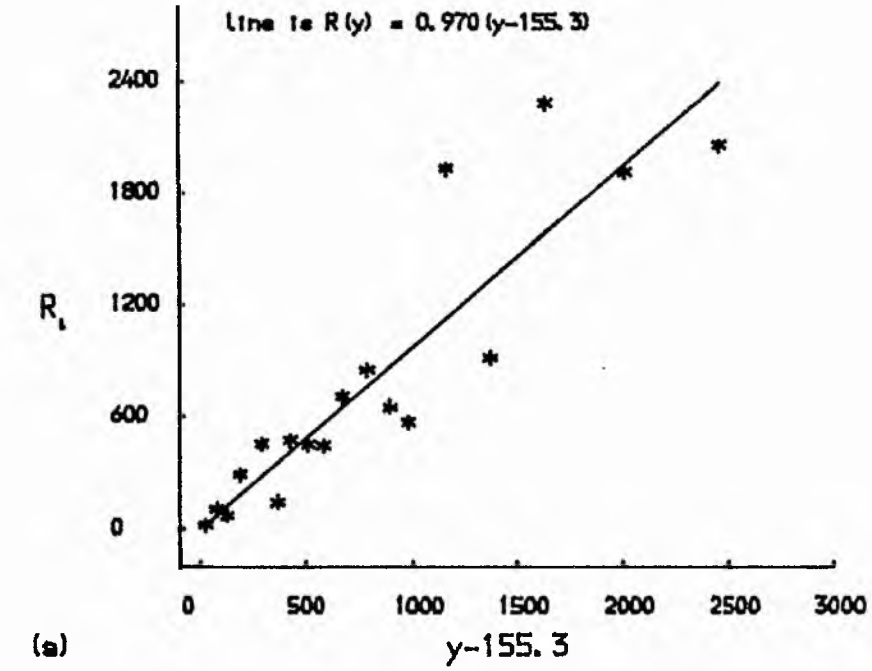


Figure 4.2. Extreme value analysis of Grubb's Data

(a) Estimated ratio plot

(b) Estimated log-ratio plot



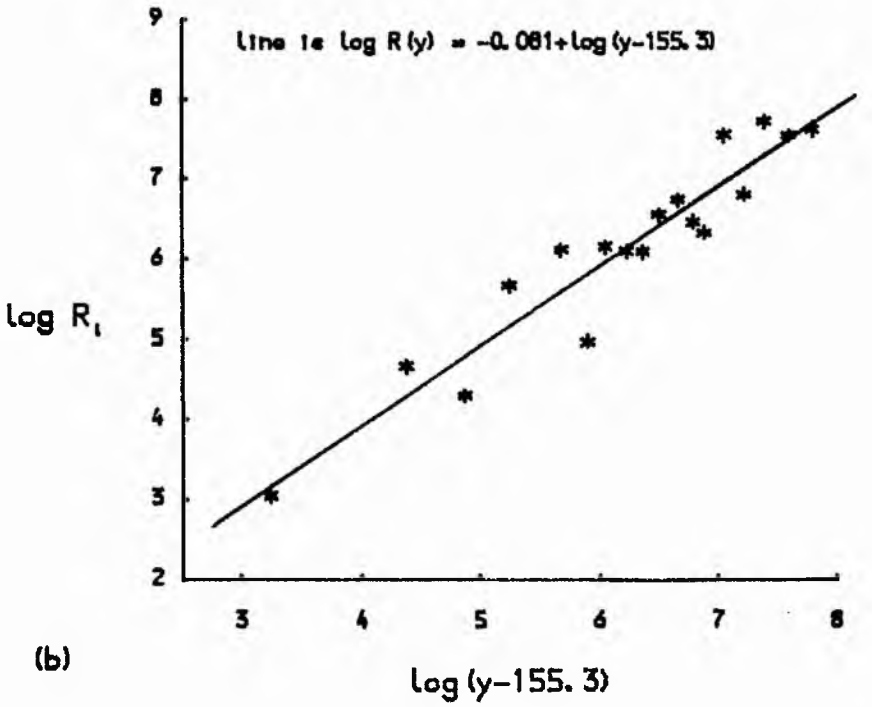


Figure 4.3. Extreme value analysis of Grubb's Data minus ratio estimate of threshold

(a) Estimated ratio plot

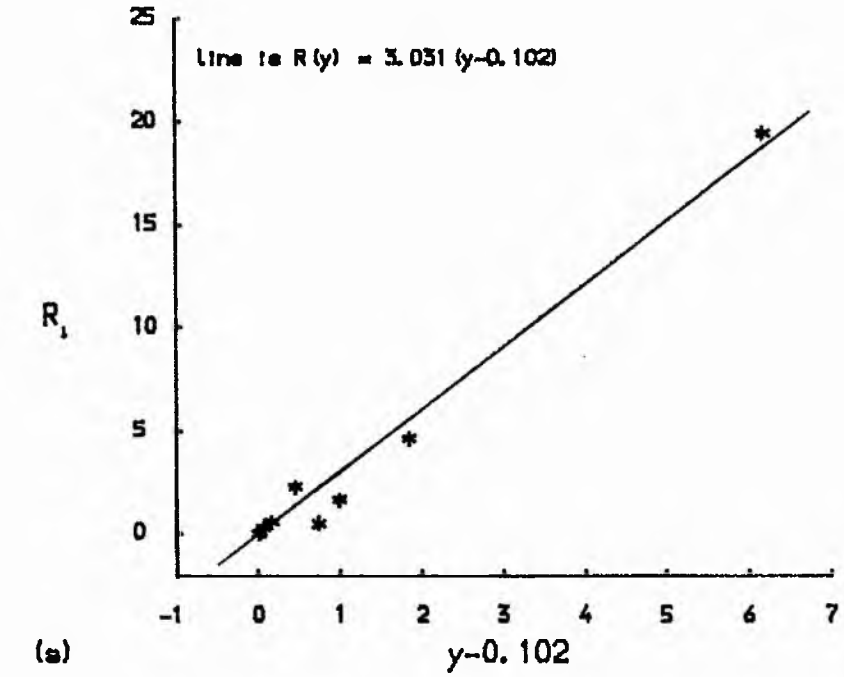
(b) Estimated log-ratio plot .

Example 4.3 Hydrogen Fluoride Data: Leidel, Busch and Lynch (1977) give a sample of twelve readings of hydrogen fluoride. The readings were 0.11, 0.11, 0.12, 0.14, 0.14, 0.21, 0.33, 0.80, 0.91, 1.30, 2.60, 10.00. D'Agostino (1986) investigated the data using a lognormal probability plot. He found that there was evidence for a non-zero threshold value which he roughly estimated to be 0.1. His probability plot estimates of the normal mean μ and standard deviation σ are -1.703 and 2.278 respectively.

The ratio plot (Figure 4.4a) for the normal analysis is roughly linear which confirms that the lognormal distribution is reasonable. The R values are averaged in the case of repeated observations. The line fitted to the ratio plot has a non-zero intercept. This together with a non-zero value of -0.437 for the transformation parameter estimate confirm that a non-zero threshold value exists. The ratio estimate for the threshold parameter is calculated using the iterative procedure described for the Weibull distribution in Section 4.4. The estimate is γ_R = 0.102.

The ratio and log-ratio plots for the data minus this ratio estimate of the threshold parameter are shown in Figure 4.4. The line fitted to the ratio plot has slope 3.031 giving σ_R = 3.031 as an estimate of σ. Expression (4.4.5) gives the estimate α_R = 0.174 which in turn gives μ_R = log(0.174) = -1.745.

The modified maximum likelihood estimates of Cohen and Whitten (1982) are calculated to be $\hat{\gamma}$ = 0.104, $\hat{\mu}$ = -1.848 and $\hat{\sigma}$ = 2.351. The parameters are estimated by simultaneously solving three equations for γ, α and β and then transforming the estimates



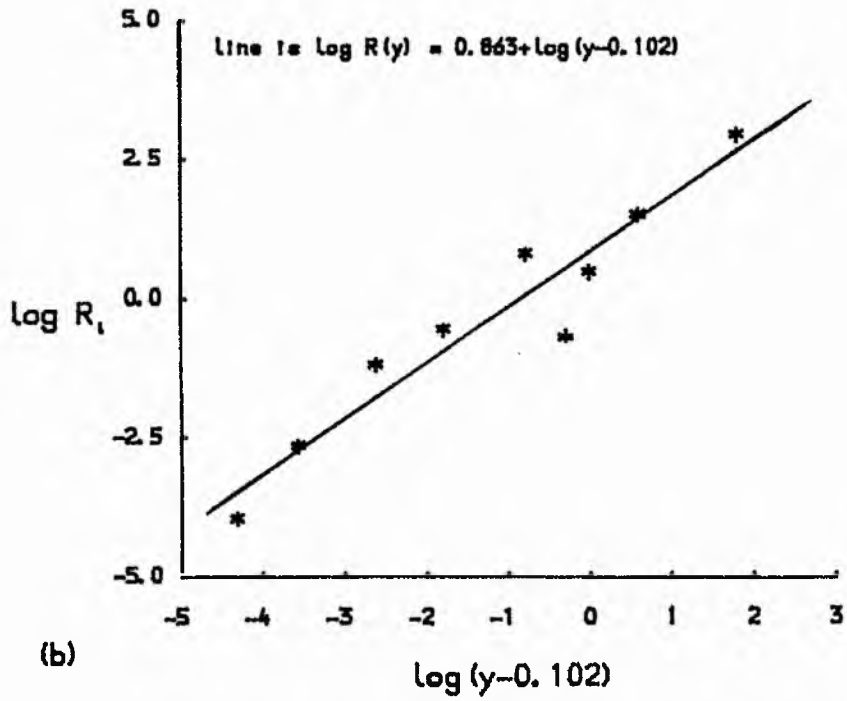


Figure 4.4. Normal analysis of Hydrogen Fluoride Data minus ratio estimate of threshold

(a) Estimated ratio plot

(b) Estimated log-ratio plot

of α and β to obtain the estimates for μ and σ. The first equation is found by equating $\log(y_i-\gamma)$ to its expected value. The other two equations are found by equating to zero the partial derivatives of the likelihood function with respect to α and β.

As in example 4.2, the goodness-of-fit is tested using the statistics W^2 and A^2 for the normal case [Stephens (1974)]. For the ratio estimation procedure the values of the test statistics are calculated to be $W^2 = 0.040$ and $A^2 = 0.378$. When the parameters are estimated using modified maximum likelihood estimation the test statistics take the values $W^2 = 0.044$ and $A^2 = 0.314$.

To conclude this example, it is noted that the ratio and log-ratio plots of Figure 4.4 suggest the presence of two parallel straight lines. This indicates that there may be two components in the data and that the sample may be drawn from two separate lognormal distributions.

4.7 DISCUSSION

Tarter and Kowalski defined the ratio R(y) to test for normality. Where normality is rejected they rely on the estimated plots of R(y) to suggest transformations to normality. In this chapter the definition of the ratio R(y) is generalised to provide a test for any distribution (with associated random variable Y) for which the following holds: some power transformation of Y has a distribution which does not allow for an unknown shape

parameter. Where the hypothesised distribution does not include an unknown shape parameter and use of the ratio R(y) leads to rejection of the hypothesised distribution, considering a family of power transformations facilitates estimation of the transformation to the distribution.

Application to examples illustates how the plots of R(y) can discover phenomena such as influential observations, outliers and contamination. The plotting methods can be used to provide parameter estimates in many models. A procedure for estimating the threshold parameter in models including the three-parameter Weibull and lognormal distributions, is developed. Although the plots may sometimes show discouraging variability, a simulation study indicates that for the case of the three-parameter Weibull distribution, the proposed estimators perform better than the modified maximum likelihood estimators of Cohen and Whitten (1982) as regards bias and root mean squared error.

The ratio-concept may prove useful in a number of other applications: e.g. identifying other models including the logistic and log-logistic distributions; the development of goodness-of-fit tests based on ratio plots; the comparison of distributions and the development of useful distributions given the functional form of the ratio R(y).

CHAPTER 5

DISCUSSION AND RECOMMENDATIONS FOR FURTHER WORK

5.1 TRANSFORMATION AND MODEL BUILDING

The purpose of this chapter is to review the methods of earlier chapters and to suggest some further developments. The major theme has been the use of transformation in model building and the analysis of data. For the purposes of this thesis the situations in which a transformation of data might prove worthwhile can be conveniently arranged in two classes. In the first, the expected responses are related to explanatory variables by some function of unknown parameters. A transformation is selected to give a linear model with constant error variance, an additive structure and a proposed error distribution.

In the second class, the responses are independent. A distribution is proposed for the observations but there is evidence to believe that the proposed distribution is not the true underlying distribution. A transformation is selected so that the distribution of the transformed responses is close to the proposed distribution. The class of distributions under consideration is the family of distributions which are not dependent on an unknown shape parameter. This family includes the gamma distribution (known order) as well as the normal, exponential, Gumbel and logistic distributions.

We proceed in both situations by specifying a family of

transformations indexed by a parameter λ and then use the data to choose a transformation that may result in a model with the desirable properties. Other parameters of interest are simultaneously estimated. The family of transformations is the power transformation family

$$y^{(\lambda)} = \begin{cases} y^{\lambda}, & \lambda \neq 0 \\ \log y, & \lambda = 0, \end{cases} \qquad (5.1.1)$$

which was redefined by Box and Cox (1964) to assure continuity at $\lambda = 0$

$$y^{(\lambda)} = \begin{cases} (y^{\lambda} - 1)/\lambda, & \lambda \neq 0 \\ \log y, & \lambda = 0. \end{cases} \qquad (5.1.2)$$

These families of power transformations can be applied in any problem with positive responses.

5.2 TRANSFORMATION AND THE LAPLACE DISTRIBUTION IN REGRESSION

The usual distributional assumption behind the linear model is that the errors are normally distributed. When too many large errors occur the normal assumption is replaced by the assumption that the errors follow the Laplace distribution, which is symmetric about zero and has tails that are heavier than those associated with the normal distribution. The method for estimating the transformation parameter is based on likelihood

considerations and is presented in Chapter 2, together with some background information and a more detailed description of the motivation for considering Laplace errors.

With the assumption of Laplace errors, the likelihood procedure for selecting a transformation is based on the criterion of minimising the sum of absolute errors. The procedure is invariant under rescaling of the response if the design matrix contains a vector of ones. Diagnostic methods are proposed for detecting the influence of individual observations on a choice of transformation. The diagnostics are based on the slope of the log-likelihood of the observations at the hypothesised value of the transformation parameter. Chapter 2 also introduces a test procedure for detecting outliers in simple linear minimum sum of absolute errors (MSAE) regression. The test statistic is the largest standardised residual and percentage points were obtained by simulation.

These procedures are robust against outlying observations. The MSAE approach and the procedures are exemplified by applying them to well-known examples including the "salinity data" (Ruppert and Carroll, 1980) and the "stack loss data" of Brownlee (1965). The approach reconciled observations, outlying under the normality assumption, to the MSAE model and the diagnostic methods furnished valuable information on influence.

Likelihood estimation under the assumption of Laplace errors, which leads naturally to MSAE estimation, is designed to handle outliers in errors but cannot deal with influential

observations caused by high leverage. Ruppert and Carroll (1985) proposed estimators for the transformation parameter that can handle outliers in both the residual and the explanatory variables when the errors are normally distributed. Similarly robust estimators are required when the errors are Laplace distributed.

Another area worthy of future investigation is the effects on the regression output resulting from transformation of the response. Differentiating the MSAE estimates may provide a measure of local sensitivity to the transformation parameter. Lawrance (1987) constructed diagnostic methods for the transformation parameter under the normality assumption, from the local changes to the parameter estimates caused by perturbing assumptions of the model. The diagnostics are useful in detecting groups of influential cases, thereby indicating possible masking effects. The development of similar diagnostics from local influence when the errors are drawn from the Laplace distribution, offers a further area of research.

5.3 TRANSFORMATION AND THE GAMMA DISTRIBUTION

The gamma distribution is one of the most important parametric models in the analysis of lifetime data. Chapter 3 considers the situation where the responses are independent and the distributional assumption that the responses follow the gamma distribution (known order) is not satisfied. Instead, the errors follow some other known distribution. An information number

approach is adopted for selecting the transformation to the gamma distribution and is based on the known distribution of the response variable. The Kullback-Leibler information number is used as a measure of discrepancy between two distributions. The transformation parameter and the gamma scale parameter are chosen to minimise the information number between the true density of the transformed variable and the density of the gamma distribution (known order). The approach yields estimates which are the limiting values of the maximum likelihood estimates (Draper and Guttman, 1968).

The information number approach is illustrated for the lognormal, Weibull and Pareto families of pdf's. The improvement towards the gamma distribution introduced by transformation is assessed numerically and graphically. The improvement is substantial when the underlying distribution is Weibull or lognormal, but transforming is not successful for the Pareto distribution. The information number approach has the additional advantage that it determines the lognormal and Weibull distributions that are closest to a gamma distribution, as measured by Kullback-Leibler divergence.

The information number approach has produced some useful results for transforming to the gamma distribution. The search for alternative measures of discrepancy to Kullback-Leibler information number might prove fruitful and still more informative.

The gamma distribution is the central member of the

Pearson family of distributions. We could also explore the merit of transforming to the central members of other families of distributions, like the Johnson or Burr families (Ord, 1972).

5.4 TRANSFORMATION, MODEL TESTING AND ESTIMATION

A graphical technique is proposed in Chapter 4 for estimating transformations of the independent responses so that the transformed data follow distributions which are not dependent on an unknown shape parameter. The technique is extended to include model testing and estimation for these distributions and for any proposed distribution which with the aid of a power transformation can be put in the simple form of a distribution dependent only on unknown threshold and scale parameters. The technique is based on a ratio which is constructed from the power transformation. The properties of the ratio for a hypothesised distribution may be used to judge the correctness of the distribution and to supply estimates of the parameters of the distribution when it appears reasonable. The test for the distribution usually takes the form: the plot of the ratio for the data will be a straight-line plot if the hypothesised distribution is the true underlying distribution. This is a useful property as departure from a straight line is easy to judge.

For the case of transforming to the uniform distribution, the ratio is the reciprocal of the density function of the

underlying distribution. For the case of transforming to the exponential distribution, the ratio is the reciprocal of the failure rate of the underlying distribution. The generalization of the ratio for transforming to a distribution which is not dependent on an unknown shape parameter, takes the form of the reciprocal of the generalised failure rate (Barlow and Van Zwet, 1969).

Transforming to the exponential, Gumbel and normal distributions is treated in detail in Chapter 4, leading to the development of model testing and estimation procedures for not only these distributions, but the uniform, Pareto, Weibull and lognormal distributions also. The procedures apply identically to the Weibull and lognormal distributions because the Weibull distribution is transformed to the Gumbel distribution in the same way that the lognormal distribution is transformed to the normal distribution. The ratio for both distributions is constructed from the log transformation.

The main strength of the ratio concept is its generality and applicability to testing for and estimating a variety of distributions. Another important advantage is that it can be used effectively with singly Type 1 and Type 2 censored data. The iterative procedure for estimating the threshold parameter of the Weibull and lognormal distributions, which is described in Chapter 4, is a further development of the ratio concept. Also presented are the results of a simulation study conducted to compare the performances of ratio estimation and modified maximum likelihood estimation (Cohen and Whitten, 1982) for the three-parameter

Weibull distribution and small to moderate sized samples. Ratio estimation has four definite advantages over modified maximum likelihood estimation. First, it does not involve numerically solving nonlinear equations. Second, estimates are always found and third, the simulation study shows that the ratio estimators perform better as far as bias and root mean squared error are concerned. Finally, ratio plotting is a powerful tool for screening data for outliers, influential observations and the presence of mixtures (or contamination). Outliers and influential observations appear as observations separated from the rest of the sample. Underlying mixtures of distributions surface as straight line segments in the ratio plot. The use of the ratio plotting technique for checking distributions, detecting outliers, influential observations and contamination and for supplying estimates within a model, is illustrated for real data in Chapter 4.

The ratio concept offers several interesting areas of research which invite future investigation. These are introduced briefly here.

The log-logistic distribution is related via the log transformation, to the logistic distribution. This suggests that the model testing and estimation procedure proposed for the Weibull and lognormal distributions, is immediately applicable to the log-logistic distribution. The performance of the procedure for the log-logistic distribution needs to be assessed and the

and the search for other distributions that lend themselves to the ratio concept could prove worthwhile.

The application of the ratio concept to the Gumbel distribution was discussed in Chapter 4. The Gumbel distribution is a special case of the generalised extreme value distribution whose distribution function may be written as

$$F(y) = \begin{cases} 1 - \exp[-\{1 + k(y-\nu)/\theta\}^{1/k}] & (k \neq 0) \\ 1 - \exp[-\exp\{(y-\nu)/\theta\}] & (k = 0), \end{cases} \qquad (5.4.1)$$

where k is a shape parameter. The case k = 0 corresponds to the Gumbel or Fisher-Tippett type I extreme value distribution while the cases k > 0 and k < 0 correspond to the Fisher-Tippett types II and III respectively. The generalised extreme value distribution is related to the Gumbel distribution by the easily shown fact that if Y has a generalised extreme value distribution with cumulative (5.4.1), then

$$\psi(Y) = k^{-1} \log [1 + k(y-\nu)/\theta] \qquad (5.4.2)$$

has a standard Gumbel distribution.

It is often of interest in hydrology to test whether a sample of data follow a Gumbel rather than a generalised extreme value distribution. This is equivalent to testing whether k = 0 in the generalised extreme value distribution. The ratio for transforming the generalised extreme value distribution to the Gumbel distribution is of the form

$$R(y) = \theta + k(y-\nu). \qquad (5.4.3)$$

The shape parameter k can be estimated as the slope of the line fitted to the ratio plot. A horizontal band of points is

indicative of the Gumbel distribution. The ratio plot can also identify observations that are influencing the selection of the shape parameter.

This argument could be developed further, resulting perhaps in a formal numerical test of the hypothesis that $k = 0$. Power comparisons could then be made with competing test procedures (Hosking, 1984).

The whole subject of hypothesis testing and goodness-of-fit could be explored further for the ratio concept. Many of the graphical tests described for hypothesised distributions relied on straight-line plots with departure from a fitted straight line measuring the discrepancy from the hypothesised distribution. A formal numerical test of goodness-of-fit for the hypothesised distribution could perhaps be based on the sum of squares of the residuals for the fitted line. Unequal variances and serial correlation are complicating factors. This and other possible approaches based on the ratio concept, may be worthy of further investigation.

The comparison of distributions is another area of hypothesis testing where the ratio concept could be applied. When the distributions are three-parameter Weibull distributions the main interest is in comparing shape and threshold parameters. Consider random samples from m Weibull distributions. The first step is to estimate ratio plots for the m samples and fit regression lines to the plots. Testing for a common shape parameter is now equivalent to testing for parallel regression lines. Testing for a common threshold parameter is equivalent to

testing whether the regression lines have equal intercepts on the horizontal axis, but slopes are arbitrary. Testing for no sample difference in the Weibull distributions is equivalent to testing for equal regression lines.

This argument could perhaps be used as a basis for devising formal test procedures for comparison of Weibull distributions, with immediate application to the lognormal and log-logistic distributions. The argument could be extended to testing whether generalised extreme value distributions have a common shape parameter and modified for the comparison of a variety of other distributions.

The hazard function or failure rate is a very useful function in life testing, since it describes the way in which the instantaneous probability of death for an individual changes with time. The failure rate is defined as

$$h(y) = \frac{f(y)}{1 - F(y)}. \qquad (5.4.4)$$

It was explained in Chapter 4 that for the case of transforming to the exponential distribution the ratio is the reciprocal failure rate and the ratio plot can be assessed as the inverted plot of the failure rate. Because of this property, the ratio plot could be used to graphically test for a constant failure rate against the alternative of a failure rate involving a change point, defined as

$$h(y) = \begin{cases} \zeta, & y \leq \xi \\ \rho\zeta, & y > \xi. \end{cases} \qquad (5.4.5)$$

This function has three unknown parameters which could be estimated from the ratio plot. This has immediate application in the field of leukemia research, where researchers are interested in determining whether a new therapy causes a departure from a constant relapse rate after remission induction. The departure is that for which a considerable reduction in relapse rate occurs after a constant relapse rate has been evident for some period after induction.

It follows from the definition of the failure rate given in (5.4.4) that the probability density function f(y) and the distribution function F(y) can be determined from the functional form of the failure rate. The conditions $F(0^-) = 0$ and $F(+\infty) = 1$ are assumed. We have from (5.4.4) that

$$h(y)dy = \frac{dF(y)}{1-F(y)}$$

or

$$\int_0^t h(y)dy = -\log[1 - F(y)]\Big|_0^t.$$

Thus,

$$\log\left[\frac{1 - F(t)}{1 - F(0)}\right] = -\int_0^t h(y)dy,$$

giving

$$1 - F(t) = \exp\left[-\int_0^t h(y)dy\right]. \qquad (5.4.6)$$

Taking derivatives, we find that

$$f(t) = h(t) \exp\left[-\int_0^t h(y)dy\right]. \qquad (5.4.7)$$

This technique has been used to develop failure distributions. The generalisation of the ratio for transforming to a distribution which does not depend on an unknown shape parameter, is the reciprocal of the generalised failure rate. It follows that new distributions could perhaps be developed from consideration of the functional form of the ratio for transforming to distributions other than the exponential distribution.

5.5 FINAL REMARKS

The transformation methods of this work have produced some interesting results and useful applications. Various extensions of these methods and recommendations for further work have been suggested. Some preliminary aspects of further work have been examined but further research is warranted. Some of the extensions and areas which require further investigation appear to be quite straightforward but are by no means complete. Other elements are considerably less obvious and provide challenging areas of research.

APPENDIX A

THE STABILIZED PROBABILITY PLOT

Let $t_1 \leq \ldots \leq t_n$ be an ordered random sample with underlying distribution function F and suppose a continuous location-scale distribution $F_o\{(t-\mu)/\sigma\}$ is hypothesised. Then the stabilised probability plot is formed by plotting

$$v_i = (2/\pi) \arcsin[F_o^{i/2}\{(t_i-\mu)/\sigma\}] \qquad (A.1)$$

against

$$u_i = (2/\pi) \arcsin[p_i^{1/2}] \qquad (A.2)$$

for $i=1,\ldots,n$, where p_i is the plotting position. For the hypothesis of normality the plotting position is $p_i = (i-3/8)/(n+1/4)$ and $p_i = (i-1/2)/n$ is chosen for use with a proposed exponential model. The parameters μ and σ in (A.1) are replaced by their maximum likelihood estimates.

APPENDIX B

MINIMISING THE INFORMATION NUMBER

The information number between the true density of the transformed variable Z and the gamma density is

$$I[f_\lambda, g_{m,\theta}] = \int f_\lambda(z) \log \left[\frac{f_\lambda(z)}{g_{m,\theta}(z)}\right] dz. \qquad (B.1)$$

This can be re-expressed as

$$E_f\left\{\log[f_\lambda(z)]\right\} - E_f\left\{\log[g_{m,\theta}(z)]\right\}, \qquad (B.2)$$

which can be expanded to

$$\begin{cases} E_f\left\{\log[f_\lambda(z)]\right\} - \lambda(m-1)E_f(\log y) + \frac{1}{\theta} E_f(y^\lambda) + \log[\Gamma(m)] \\ \qquad + m\log\theta, \qquad \lambda \neq 0 \\ E_f\left\{\log[f_\lambda(z)]\right\} - (m-1)E_f[\log(\log y)] + \frac{1}{\theta} E_f(\log y) + \log[\Gamma(m)] \\ \qquad + m\log\theta, \qquad \lambda = 0, \end{cases} \qquad (B.3)$$

giving

$$\begin{aligned} &E_f\left\{\log[f(y)]\right\} - (\lambda-1)E_f(\log y) - \log\lambda - \lambda(m-1)E_f(\log y) \\ &\qquad + \frac{1}{\theta} E_f(y^\lambda) + \log[\Gamma(m)] + m\log\theta, \quad \lambda \neq 0 \\ &E_f\left\{\log[f(y)]\right\} + E_f(\log y) - (m-1)E_f[\log(\log y)] + \frac{1}{\theta} E_f(\log y) \\ &\qquad + \log[\Gamma(m)] + m\log\theta, \quad \lambda = 0. \end{aligned} \qquad (B.4)$$

For λ fixed, the minimising value of θ is

$$\theta^*(\lambda) = \begin{cases} \frac{1}{m} E_f(y^\lambda), & \lambda \neq 0 \\ \frac{1}{m} E_f(\log y), & \lambda = 0. \end{cases} \tag{B.5}$$

Substitution of this expression for $\theta^*(\lambda)$ in (B.4) gives the information number as a function of λ

$$K(\lambda) = \begin{cases} E_f\{\log[f(y)]\} + (1-m\lambda)E_f(\log y) - \log\lambda + m \\ \quad + \log[\Gamma(m)] - m\log m + m\log[E_f(y^\lambda)], \quad \lambda \neq 0 \\ \\ E_f\{\log[f(y)]\} + E_f(\log y) - (m-1)E_f[\log(\log y)] \\ \quad + m + \log[\Gamma(m)] - m\log m + m\log[E_f(\log y)], \quad \lambda = 0. \end{cases} \tag{B.6}$$

APPENDIX C

ESTIMATION OF THE RATIO PLOT

Consider the ordered sample $y_1 < y_2 < \ldots < y_n$ and the sample cumulative of the form

$$F^*(y_i) = p_i. \tag{C.1}$$

At each y_i, F^* has a jump of size δ and may be smoothed by connecting the points $F^*(y_i)$ by straight lines. An estimate of $fF^{-1}(t)$ is provided by the slope of the line connecting $F^*(y_{[t/\delta]})$ and $F^*(y_{[t/\delta]+1})$, where [x] is the largest integer less than x. The slope is

$$\delta\left\{y_{[t/\delta]+1} - y_{[t/\delta]}\right\}^{-1} \tag{C.2}$$

and hence the graph of R(y) may be estimated by plotting $(y_{i+1} - y_i)\phi\Phi^{-1}(p_i)/\delta$ against $(y_i + y_{i+1})/2$ for $i = 1,2,\ldots,n-1$.

REFERENCES

ABRAMOWITZ, M & STEGUN, I A, Eds (1965). Handbook of Mathematical Functions. New York : Dover.

ANDREWS, D F (1974). A robust method for multiple linear regression. Technometrics 16, 523–531.

ATKINSON, A C (1982). Regression diagnostics, transformations and constructed variables (with discussion). J. Roy. Statist. Soc. B44, 1–36.

ATKINSON, A C (1983). Diagnostic regression analysis and shifted power transformations. Technometrics 25, 23–33.

ATKINSON, A C (1985). Plots, Transformations and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis. Oxford : Clarendon.

BARLOW, R E & VAN ZWET, W R (1969). Asymptotic properties of isotonic estimators for the generalised failure rate function, Part II: Asymptotic distribution. Research Report ORC 69–10, University of California, Berkeley.

BARRODALE, I & ROBERTS, F D K (1973). An improved algorithm for discrete L_1 linear approximation. Siam J. Numer. Anal. 10, 839–848.

BLOCH, D A & GASTWIRTH, J L (1968). On a simple estimate of the reciprocal of the density function. Ann. Math. Statist. 39, 1083–1085.

BLOM, G (1958). Statistical Estimates and Transformed Beta–Variables. New York : Wiley.

BLOOMFIELD, P & STEIGER, W L (1983). Least Absolute Deviations: Theory, Applications and Algorithms. Boston : Birkhauser.

BOX, G E P & COX, D R (1964). An analysis of transformations (with discussion). J. Roy. Statist. Soc. B26, 211-252.

BROWNLEE, K A (1965). Statistical Theory and Methodology in Science and Engineering (2nd ed). New York : Wiley.

BURY, K V (1975). Statistical Models in Applied Science. New York : Wiley.

CARROLL, R J (1980). A robust method for testing transformations to achieve approximate normality. J. Roy. Statist. Soc. B42, 71-78.

CARROLL R J (1982). Two examples of transformations when there are possible outliers. Applied Statistics 31, 149-152.

CHAPMAN, H M & DEMERITT, D B (1936). Elements of Forest Mensuration (2nd ed) Albany, New York : Williams Press.

CHARNES, A, COOPER, W W & FERGUSON, R O (1955). Optimal estimation of executive compensation by linear programming. Management Sci. 1, 138-151.

COHEN, A C & WHITTEN, B J (1980). Estimation in the three-parameter lognormal distribution. J. Am. Statist. Assoc. 75, 399-404.

COHEN, A C & WHITTEN, B J (1982). Modified maximum likelihood and modified moment estimators for the three-parameter Weibull distribution. Comm. Statist. A11, 2631-2656.

COOK, R D & WANG, P C (1983). Transformations and influential cases in regression. Technometrics 25, 337-343.

COX, D R & HINKLEY, D V (1978). Problems and Solutions in Theoretical Statistics. London : Chapman & Hall.

D'AGOSTINO, R B (1986). Graphical Analysis. Goodness-of-Fit Techniques (eds R B D'Agostino & M A Stephens), pp 7–62. New York : Dekker.

DRAPER, N R & GUTTMAN, I (1968). Transformations of life-test data. Canad. Math. Bull. 11, 475–488.

FISHER, R A & TIPPETT, L H C (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. Proc. Camb. Phil. Soc. 24, 180–190.

GAVIN, W W (1960). Introduction to Linear Programming. New York : McGraw-Hill.

GRINGORTEN, I I (1963). A plotting rule for extreme probability paper. J. Geophysical Research 68, 813–814.

GRUBBS, F E (1971). Approximate fiducial bounds on reliability for the two parameter negative exponential distribution. Technometrics 13, 873–876.

HAMPEL, F R, RONCHETTI, E M, ROUSSEEUW, P J & STAHEL, W A (1986). Robust Statistics: The Approach Based on Influence Functions. New York : Wiley.

HERNANDEZ, F & JOHNSON, R A (1980). The large-sample behaviour of transformations to normality. J. Am. Statist. Assoc. 75, 855–861.

HOSKING, J R M (1984). Testing whether the shape parameter is zero in the generalized extreme-value distribution. Biometrika 71, 367–374.

HUBER, P J (1972). Robust Statistics: A Review. Ann. Math.

Statist. 43, 1041-1067.

JONES, R L & ROWCLIFFE, D J (1979). Tensile-strength distributions for silicon nitride and silicon carbide ceramics. Ceramic Bulletin 59, 836-839 and 844.

KAPPENMAN, R F (1985). Estimation for the three-parameter Weibull, lognormal and gamma distributions. Computational Statistics and Data Analysis 3, 11-23.

KRASKER, W S & WELSCH, R E (1982). Efficient bounded-influence regression equations. J. Am. Statist. Assoc. 77, 595-604.

KULLBACK, S (1968). Information Theory and Statistics. New York : Dover Publications.

LAWLESS, J F (1977). Prediction intervals for the two parameter exponential distribution. Technometrics 19, 469-472.

LAWRANCE, A J (1987). Regression transformation diagnostics using local influence (preprint).

LEIDEL, N A, BUSCH, K A & LYNCH, J R (1977). Occupational Exposure Sampling Strategy Manual. U.S. Dept. of H.E.W., Cincinnati, Ohio.

MICHAEL, J R (1983). The stabilised probability plot. Biometrika 70, 11-17.

MUNRO, A H & WIXLEY, R A (1970). Estimators based on order statistics of small samples from a three-parameter lognormal distribution. J. Am. Statist. Assoc. 65, 212-225.

ORD, J K (1972). Families of Frequency Distributions. London : Griffin.

RUPPERT, D & CARROLL, R J (19980). Trimmed least squares estimation in the linear model. J. Am. Statist. Assoc. 75,

828–838.

RUPPERT, D & CARROLL, R J (1985). Transformations in regression: a robust analysis. Technometrics 27, 1–12.

SCHLESSELMAN, J (1971). Power families: a note on the Box and Cox transformation. J. Roy. Statist. Soc. B33, 207–211.

SIDDIQUI, M M (1960). Distribution of quantiles in samples from a bivariate population. J. Res. N.B.S. 64B, 145–150.

SNEDECOR, G W & COCHRAN, W G (1967). Statistical Methods (6th ed). Ames, Iowa : Iowa State University Press.

SPOSITO, V A, SMITH, W C & McCORMICK, G (1978). Minimising the Sum of Absolute Deviations. Gottingen, W. Germany : Vandenhoeck and Ruprecht.

STEPHENS, M A (1974). EDF statistics for goodness of fit and some comparisons. J. Am. Statist. Assoc. 69, 730–737.

STEPHENS, M A (1977). Goodness of fit for the extreme value distribution. Biometrika 64, 583–588.

SUTCLIFFE, J V et al (1975). Flood Studies Report, Volume I: Hydrological Studies. National Environmental Research Council, London.

TARTER, M E & KOWALSKI, C J (1972). A new test for and class of transformations to normality. Technometrics 14, 735–744.

TIETJEN, G L, MOORE, R H & BECKMAN, R J (1973). Testing for a single outlier in simple linear regression. Technometrics 15, 717–721.

WEISBERG, S (1980). Applied Linear Regression. New York : Wiley.